

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION  
International Bureau



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification 6 :</b> C12N 15/12, 5/10, 1/21, C07K 14/47, 16/18, C12N 1/21, C07K 14/47, 16/18, C12Q 1/68, G01N 33/50, 33/53, 33/68, A61K 38/17		<b>A2</b>	<b>(11) International Publication Number:</b> WO 98/39448 <b>(43) International Publication Date:</b> 11 September 1998 (11.09.98)																														
<b>(21) International Application Number:</b> PCT/US98/04493 <b>(22) International Filing Date:</b> 6 March 1998 (06.03.98) <b>(30) Priority Data:</b> <table border="0"><tr><td>60/040,162</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,333</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/038,621</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,161</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,626</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,334</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,336</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/040,163</td><td>7 March 1997 (07.03.97)</td><td>US</td></tr><tr><td>60/043,580</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr><tr><td>60/043,568</td><td>11 April 1997 (11.04.97)</td><td>US</td></tr></table> <p><i>(Continued on the following page)</i></p> <b>(71) Applicant (for all designated States except US):</b> HUMAN GENOME SCIENCES, INC. [US/US]; 9410 Key West Avenue, Rockville, MD 20850 (US). <b>(72) Inventors; and</b> <b>(75) Inventors/Applicants (for US only):</b> RUBEN, Steven, M. [US/US]; 18528 Heritage Hills Drive, Olney, MD 20832 (US). ROSEN, Craig, A. [US/US]; 22400 Rolling Hills Road, Laytonsville, MD 20882 (US). FISCHER, Carrie, L. [US/US]; 5810 Hall Street, Burke, VA 22015 (US). SOPER, Daniel, R. [US/US]; 15050 Stillfield, Place, Centreville, VA 22020 (US). CARTER, Kenneth, C. [US/US]; 11601 Brandy Hall Lane, North Potomac, MD 20878 (US). BEDNARIK, Daniel, P. [US/US]; 8822 Blue Sea Drive, Columbia, MD 21046 (US). ENDRESS, Gregory, A. [US/US]; 9729 Clagett Farm Drive, Potomac, MD 20854 (US). YU, Guo-Liang [CN/US]; 13524 Straw Bale Lane, Darnestown, MD 20878 (US). NI, Jian [CN/US]; 5502 Manorfield Road, Rockville, MD 20853 (US). FENG, Ping [CN/US]; 4 Relda Court, Gaithersburg, MD 20878 (US). YOUNG, Paul, E. [US/US]; 122 Beckwith Street, Gaithersburg, MD 20878 (US). GREENE, John, M. [US/US]; 872 Diamond Drive, Gaithersburg, MD 20878 (US). FERRIE, Ann, M. [US/US]; 13203 L Astoria Hill Court, Germantown,		60/040,162	7 March 1997 (07.03.97)	US	60/040,333	7 March 1997 (07.03.97)	US	60/038,621	7 March 1997 (07.03.97)	US	60/040,161	7 March 1997 (07.03.97)	US	60/040,626	7 March 1997 (07.03.97)	US	60/040,334	7 March 1997 (07.03.97)	US	60/040,336	7 March 1997 (07.03.97)	US	60/040,163	7 March 1997 (07.03.97)	US	60/043,580	11 April 1997 (11.04.97)	US	60/043,568	11 April 1997 (11.04.97)	US	MD 20874 (US). DUAN, Roxanne [US/US]; 4541 Fairfield Drive, Bethesda, MD 20814 (US). HU, Jing-Shan [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). FLORENCE, Kimberly, A. [US/US]; 12805 Atlantic Avenue, Rockville, MD 20851 (US). OLSEN, Henrik, S. [DK/US]; 182 Kendrick Place #24, Gaithersburg, MD 20878 (US). EBNER, Reinhard [DE/US]; 9906 Shelburne Terrace #316, Gaithersburg, MD 20878 (US). BREWER, Laurie, A. [US/US]; 14920 Mount Nebo Road, Poolesville, MD 20837 (US). MOORE, Paul, A. [GB/US]; Apartment #104, 1908 Holly Ridge Drive, McLean, VA 22102 (US). SHI, Yanggu [CN/US]; 437 West Side Drive, Gaithersburg, MD 20878 (US). LAFLEUR, David, W. [US/US]; 1615 Q Street, N.W. #807, Washington, DC 20009 (US). LI, Yi [CN/US]; 1247 Lakeside Drive #3034, Sunnyvale, CA 94086 (US). ZENG, Zhizhen [CN/US]; 13950 Saddleview Drive, Gaithersburg, MD 20878 (US). KYAW, Hla [BU/US]; 520 Sugarbush Circle, Frederick, MD 21703 (US). <b>(74) Agents:</b> BROOKES, Anders, A. et al.; Human Genome Sciences, Inc., 9410 Key West Avenue, Rockville, MD 10850 (US). <b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, GW, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).	
60/040,162	7 March 1997 (07.03.97)	US																															
60/040,333	7 March 1997 (07.03.97)	US																															
60/038,621	7 March 1997 (07.03.97)	US																															
60/040,161	7 March 1997 (07.03.97)	US																															
60/040,626	7 March 1997 (07.03.97)	US																															
60/040,334	7 March 1997 (07.03.97)	US																															
60/040,336	7 March 1997 (07.03.97)	US																															
60/040,163	7 March 1997 (07.03.97)	US																															
60/043,580	11 April 1997 (11.04.97)	US																															
60/043,568	11 April 1997 (11.04.97)	US																															
<b>(54) Title:</b> 186 HUMAN SECRETED PROTEINS <b>(57) Abstract</b> <p>The present invention relates to 186 novel human secreted proteins and isolated nucleic acids containing the coding regions of the genes encoding such proteins. Also provided are vectors, host cells, antibodies, and recombinant methods for producing human secreted proteins. The invention further relates to diagnostic and therapeutic methods useful for diagnosing and treating disorders related to these novel human secreted proteins.</p>		<b>Published</b> Without international search report and to be republished upon receipt of that report.																															

=> s wo9839448/pn

L1 1 W09839448/PN

=> d bib

Full  
Text

OF 1 WPIDS COPYRIGHT 2004 THOMSON DERWENT on STN

AN 1998-506364 [43] WPIDS  
CR 1998-414099 [35]; 1998-414100 [35]; 1998-414105 [35]; 1998-414114 [35];  
1998-427559 [36]; 1998-520811 [44]; 1998-609887 [51]; 1999-059865 [05];  
1999-080881 [07]; 1999-120770 [10]; 1999-132229 [11]; 1999-132234 [11];  
1999-190160 [16]; 1999-204988 [17]; 1999-418749 [35]; 1999-430031 [36];  
1999-551363 [46]; 2000-106100 [09]; 2000-126931 [11]; 2000-161128 [14];  
2000-182442 [16]; 2000-195282 [17]; 2000-482826 [42]; 2000-665238 [64];  
2001-425865 [45]; 2001-625724 [72]; 2002-362489 [39]; 2002-574454 [61];  
2002-598780 [64]; 2002-599716 [64]; 2002-634796 [68]; 2002-730795 [79];  
2003-310989 [30]; 2003-466138 [44]; 2003-492322 [46]; 2003-511926 [48];  
2003-521800 [49]; 2003-531736 [50]; 2003-540138 [51]; 2003-540785 [51];  
2003-540804 [51]; 2003-567105 [53]; 2003-576674 [54]; 2003-829564 [77];  
2003-864797 [80]; 2003-898535 [82]; 2003-901099 [82]; 2004-042167 [04];  
2004-088563 [09]; 2004-131264 [13]; 2004-180094 [17]; 2004-225733 [21]  
DNN N1998-394741 DNC C1998-152795  
TI New isolated human genes and the secreted polypeptide(s) they encode -  
useful for diagnosis and treatment of e.g. cancers, neurological  
disorders, immune diseases, inflammation or blood disorders.  
DC B04 D13 D16 S03  
IN BEDNARIK, D P; BREWER, L A; CARTER, K C; DUAN, R; EBNER, R; ENDRESS, G A;  
FENG, P; FERRIE, A M; FISCHER, C L; FLORENCE, K A; GREENE, J M; HU, J; NI,  
J; OLSEN, H S; ROSEN, C A; RUBEN, S M; SOPPET, D R; YOUNG, P E; YU, G;  
DUAN, D R; KYAW, H; LAFLEUR, D W; LI, Y; MOORE, P A; SHI, Y; ZENG, Z;  
DILLON, P J; FAN, P; FLORENCE, C; FLORENCE, K; WEI, Y; YOUNG, P  
PA (HUMA-N) HUMAN GENOME SCI INC; (BREW-I) BREWER L A; (EBNE-I) EBNER R;  
(FISC-I) FISCHER C L; (KYAW-I) KYAW H; (LAFL-I) LAFLEUR D W; (LIYY-I) LI  
Y; (MOOR-I) MOORE P A; (OLSE-I) OLSEN H S; (ROSE-I) ROSEN C A; (RUBE-I)  
RUBEN S M; (SHIY-I) SHI Y; (SOPP-I) SOPPET D R; (ZENG-I) ZENG Z  
CYC 82  
PI WO 9839448 A2 19980911 (199843)\* EN 721<--  
RW: AT BE CH DE DK EA ES FI FR GB GH GM GR IE IT KE LS LU MC MW NL OA  
PT SD SE SZ UG ZW  
W: AL AM AT AU AZ BA BB BG BR BY CA CH CN CU CZ DE DK EE ES FI GB GE  
GH GM GW HU ID IL IS JP KE KG KP KR KZ LC LK LR LS LT LU LV MD MG  
MK MN MW MX NO NZ PL PT RO RU SD SE SG SI SK SL TJ TM TR TT UA UG  
US UZ VN YU ZW  
AU 9865453 A 19980922 (199908)  
AU 9891304 A 19990322 (199931)  
EP 972030 A2 20000119 (200009) EN  
R: AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE  
US 6342581 B1 20020129 (200210)  
JP 2002510192 W 20020402 (200225)  
US 2003022185 A1 20030130 (200311)  
US 2003092893 A1 20080515 (200335)  
US 2003166906 A1 20030904 (200359)  
EP 1352962 A1 20031015 (200368) EN  
R: AT BE CH DE DK ES FI FR GB GR IE IT LI LU MC NL PT SE  
ADT WO 9839448 A2 WO 1998-US4493 19980306; AU 9865453 A AU 1998-65453  
19980306; AU 9891304 A AU 1998-91304 19980903; EP 972030 A2 EP 1998-905127  
19980306, WO 1998-US4493 19980306; US 6342581 B1 Provisional US  
1997-51916P 19970708, Provisional US 1997-51918P 19970708, Provisional US  
1997-51919P 19970708, Provisional US 1997-51920P 19970708, Provisional US  
1997-51925P 19970708, Provisional US 1997-51926P 19970708, Provisional US

60/043,314	11 april 1997 (11.04.97)	US	60/047,596	23 May 1997 (23.05.97)	US	60/056,637	22 August 1997 (22.08.97)	US
60/043,569	11 april 1997 (11.04.97)	US	60/047,612	23 May 1997 (23.05.97)	US	60/056,903	22 August 1997 (22.08.97)	US
60/043,311	11 april 1997 (11.04.97)	US	60/047,632	23 May 1997 (23.05.97)	US	60/056,888	22 August 1997 (22.08.97)	US
60/043,671	11 april 1997 (11.04.97)	US	60/047,601	23 May 1997 (23.05.97)	US	60/056,879	22 August 1997 (22.08.97)	US
60/043,674	11 april 1997 (11.04.97)	US	60/047,595	23 May 1997 (23.05.97)	US	60/056,880	22 August 1997 (22.08.97)	US
60/043,669	11 april 1997 (11.04.97)	US	60/047,599	23 May 1997 (23.05.97)	US	60/056,894	22 August 1997 (22.08.97)	US
60/043,312	11 april 1997 (11.04.97)	US	60/047,588	23 May 1997 (23.05.97)	US	60/056,911	22 August 1997 (22.08.97)	US
60/043,313	11 april 1997 (11.04.97)	US	60/047,585	23 May 1997 (23.05.97)	US	60/056,636	22 August 1997 (22.08.97)	US
60/043,672	11 april 1997 (11.04.97)	US	60/047,586	23 May 1997 (23.05.97)	US	60/056,874	22 August 1997 (22.08.97)	US
60/043,578	11 april 1997 (11.04.97)	US	60/047,590	23 May 1997 (23.05.97)	US	60/056,910	22 August 1997 (22.08.97)	US
60/043,576	11 april 1997 (11.04.97)	US	60/047,594	23 May 1997 (23.05.97)	US	60/056,864	22 August 1997 (22.08.97)	US
60/043,670	11 april 1997 (11.04.97)	US	60/047,589	23 May 1997 (23.05.97)	US	60/056,631	22 August 1997 (22.08.97)	US
60/047,600	23 May 1997 (23.05.97)	US	60/047,593	23 May 1997 (23.05.97)	US	60/056,845	22 August 1997 (22.08.97)	US
60/047,615	23 May 1997 (23.05.97)	US	60/047,614	23 May 1997 (23.05.97)	US	60/056,892	22 August 1997 (22.08.97)	US
60/047,597	23 May 1997 (23.05.97)	US	60/047,501	23 May 1997 (23.05.97)	US	60/056,632	22 August 1997 (22.08.97)	US
60/047,502	23 May 1997 (23.05.97)	US	60/048,974	06 June 1997 (06.06.97)	US	60/056,664	22 August 1997 (22.08.97)	US
60/047,633	23 May 1997 (23.05.97)	US	60/048,964	06 June 1997 (06.06.97)	US	60/056,876	22 August 1997 (22.08.97)	US
60/047,583	23 May 1997 (23.05.97)	US	60/049,610	13 June 1997 (13.06.97)	US	60/056,881	22 August 1997 (22.08.97)	US
60/047,617	23 May 1997 (23.05.97)	US	60/051,926	08 July 1997 (08.07.97)	US	60/056,909	22 August 1997 (22.08.97)	US
60/047,618	23 May 1997 (23.05.97)	US	60/052,874	16 July 1997 (16.07.97)	US	60/056,875	22 August 1997 (22.08.97)	US
60/047,503	23 May 1997 (23.05.97)	US	60/055,724	18 August 1997 (18.08.97)	US	60/056,862	22 August 1997 (22.08.97)	US
60/047,592	23 May 1997 (23.05.97)	US	60/056,886	22 August 1997 (22.08.97)	US	60/056,887	22 August 1997 (22.08.97)	US
60/047,581	23 May 1997 (23.05.97)	US	60/056,877	22 August 1997 (22.08.97)	US	60/056,908	22 August 1997 (22.08.97)	US
60/047,584	23 May 1997 (23.05.97)	US	60/056,889	22 August 1997 (22.08.97)	US	60/056,884	22 August 1997 (22.08.97)	US
60/047,500	23 May 1997 (23.05.97)	US	60/056,893	22 August 1997 (22.08.97)	US	60/057,761	05 September 1997 (05.09.97)	US
60/047,587	23 May 1997 (23.05.97)	US	60/056,630	22 August 1997 (22.08.97)	US	60/057,650	05 September 1997 (05.09.97)	US
60/047,492	23 May 1997 (23.05.97)	US	60/056,878	22 August 1997 (22.08.97)	US	60/057,669	05 September 1997 (05.09.97)	US
60/047,598	23 May 1997 (23.05.97)	US	60/056,662	22 August 1997 (22.08.97)	US	60/058,785	12 September 1997 (12.09.97)	US
60/047,613	23 May 1997 (23.05.97)	US	60/056,872	22 August 1997 (22.08.97)	US	60/061,060	02 October 1997 (02.10.97)	US
60/047,582	23 May 1997 (23.05.97)	US	60/056,882	22 August 1997 (22.08.97)	US			

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav	TM	Turkmenistan
BF	Burkina Faso	GR	Greece		Republic of Macedonia	TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

## 186 Human Secreted Proteins

### *Field of the Invention*

This invention relates to newly identified polynucleotides and the polypeptides encoded by these polynucleotides, uses of such polynucleotides and polypeptides, and  
5 their production.

### *Background of the Invention*

Unlike bacterium, which exist as a single compartment surrounded by a membrane, human cells and other eucaryotes are subdivided by membranes into many functionally distinct compartments. Each membrane-bounded compartment, or  
10 organelle, contains different proteins essential for the function of the organelle. The cell uses "sorting signals," which are amino acid motifs located within the protein, to target proteins to particular cellular organelles.

One type of sorting signal, called a signal sequence, a signal peptide, or a leader sequence, directs a class of proteins to an organelle called the endoplasmic reticulum  
15 (ER). The ER separates the membrane-bounded proteins from all other types of proteins. Once localized to the ER, both groups of proteins can be further directed to another organelle called the Golgi apparatus. Here, the Golgi distributes the proteins to vesicles, including secretory vesicles, the cell membrane, lysosomes, and the other organelles.

Proteins targeted to the ER by a signal sequence can be released into the extracellular space as a secreted protein. For example, vesicles containing secreted proteins can fuse with the cell membrane and release their contents into the extracellular space - a process called exocytosis. Exocytosis can occur constitutively or after receipt of a triggering signal. In the latter case, the proteins are stored in secretory vesicles (or  
20 secretory granules) until exocytosis is triggered. Similarly, proteins residing on the cell membrane can also be secreted into the extracellular space by proteolytic cleavage of a "linker" holding the protein to the membrane.

Despite the great progress made in recent years, only a small number of genes encoding human secreted proteins have been identified. These secreted proteins include  
30 the commercially valuable human insulin, interferon, Factor VIII, human growth hormone, tissue plasminogen activator, and erythropoietin. Thus, in light of the pervasive role of secreted proteins in human physiology, a need exists for identifying and characterizing novel human secreted proteins and the genes that encode them. This knowledge will allow one to detect, to treat, and to prevent medical disorders by using  
35 secreted proteins or the genes that encode them.



### Summary of the Invention

The present invention relates to novel polynucleotides and the encoded polypeptides. Moreover, the present invention relates to vectors, host cells, antibodies, and recombinant methods for producing the polypeptides and polynucleotides. Also provided are diagnostic methods for detecting disorders related to the polypeptides, and therapeutic methods for treating such disorders. The invention further relates to screening methods for identifying binding partners of the polypeptides.

### Detailed Description

#### Definitions

The following definitions are provided to facilitate understanding of certain terms used throughout this specification.

In the present invention, "isolated" refers to material removed from its original environment (e.g., the natural environment if it is naturally occurring), and thus is altered "by the hand of man" from its natural state. For example, an isolated polynucleotide could be part of a vector or a composition of matter, or could be contained within a cell, and still be "isolated" because that vector, composition of matter, or particular cell is not the original environment of the polynucleotide.

In the present invention, a "secreted" protein refers to those proteins capable of being directed to the ER, secretory vesicles, or the extracellular space as a result of a signal sequence, as well as those proteins released into the extracellular space without necessarily containing a signal sequence. If the secreted protein is released into the extracellular space, the secreted protein can undergo extracellular processing to produce a "mature" protein. Release into the extracellular space can occur by many mechanisms, including exocytosis and proteolytic cleavage.

As used herein, a "polynucleotide" refers to a molecule having a nucleic acid sequence contained in SEQ ID NO:X or the cDNA contained within the clone deposited with the ATCC. For example, the polynucleotide can contain the nucleotide sequence of the full length cDNA sequence, including the 5' and 3' untranslated sequences, the coding region, with or without the signal sequence, the secreted protein coding region, as well as fragments, epitopes, domains, and variants of the nucleic acid sequence. Moreover, as used herein, a "polypeptide" refers to a molecule having the translated amino acid sequence generated from the polynucleotide as broadly defined.

In the present invention, the full length sequence identified as SEQ ID NO:X was often generated by overlapping sequences contained in multiple clones (contig

analysis). A representative clone containing all or most of the sequence for SEQ ID NO:X was deposited with the American Type Culture Collection ("ATCC"). As shown in Table 1, each clone is identified by a cDNA Clone ID (Identifier) and the ATCC Deposit Number. The ATCC is located at 12301 Park Lawn Drive, Rockville, Maryland 20852, USA. The ATCC deposit was made pursuant to the terms of the Budapest Treaty on the international recognition of the deposit of microorganisms for purposes of patent procedure.

A "polynucleotide" of the present invention also includes those polynucleotides capable of hybridizing, under stringent hybridization conditions, to sequences contained in SEQ ID NO:X, the complement thereof, or the cDNA contained within the clone deposited with the ATCC. "Stringent hybridization conditions" refers to an overnight incubation at 42° C in a solution comprising 50% formamide, 5x SSC (750 mM NaCl, 75 mM sodium citrate), 50 mM sodium phosphate (pH 7.6), 5x Denhardt's solution, 10% dextran sulfate, and 20 µg/ml denatured, sheared salmon sperm DNA, followed by washing the filters in 0.1x SSC at about 65°C.

Also contemplated are nucleic acid molecules that hybridize to the polynucleotides of the present invention at lower stringency hybridization conditions. Changes in the stringency of hybridization and signal detection are primarily accomplished through the manipulation of formamide concentration (lower percentages of formamide result in lowered stringency); salt conditions, or temperature. For example, lower stringency conditions include an overnight incubation at 37°C in a solution comprising 6x SSPE (20x SSPE = 3M NaCl; 0.2M NaH<sub>2</sub>PO<sub>4</sub>; 0.02M EDTA, pH 7.4), 0.5% SDS, 30% formamide, 100 µg/ml salmon sperm blocking DNA; followed by washes at 50°C with 1xSSPE, 0.1% SDS. In addition, to achieve even lower stringency, washes performed following stringent hybridization can be done at higher salt concentrations (e.g. 5x SSC).

Note that variations in the above conditions may be accomplished through the inclusion and/or substitution of alternate blocking reagents used to suppress background in hybridization experiments. Typical blocking reagents include Denhardt's reagent, BLOTTO, heparin, denatured salmon sperm DNA, and commercially available proprietary formulations. The inclusion of specific blocking reagents may require modification of the hybridization conditions described above, due to problems with compatibility.

Of course, a polynucleotide which hybridizes only to polyA+ sequences (such as any 3' terminal polyA+ tract of a cDNA shown in the sequence listing), or to a

complementary stretch of T (or U) residues, would not be included in the definition of "polynucleotide," since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-stranded cDNA clone).

The polynucleotide of the present invention can be composed of any polyribonucleotide or polydeoxynucleotide, which may be unmodified RNA or DNA or modified RNA or DNA. For example, polynucleotides can be composed of single- and double-stranded DNA, DNA that is a mixture of single- and double-stranded regions, single- and double-stranded RNA, and RNA that is mixture of single- and double-stranded regions, hybrid molecules comprising DNA and RNA that may be single-stranded or, more typically, double-stranded or a mixture of single- and double-stranded regions. In addition, the polynucleotide can be composed of triple-stranded regions comprising RNA or DNA or both RNA and DNA. A polynucleotide may also contain one or more modified bases or DNA or RNA backbones modified for stability or for other reasons. "Modified" bases include, for example, tritylated bases and unusual bases such as inosine. A variety of modifications can be made to DNA and RNA, thus, "polynucleotide" embraces chemically, enzymatically, or metabolically modified forms.

The polypeptide of the present invention can be composed of amino acids joined to each other by peptide bonds or modified peptide bonds, i.e., peptide isosteres, and may contain amino acids other than the 20 gene-encoded amino acids. The polypeptides may be modified by either natural processes, such as posttranslational processing, or by chemical modification techniques which are well known in the art. Such modifications are well described in basic texts and in more detailed monographs, as well as in a voluminous research literature. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. It will be appreciated that the same type of modification may be present in the same or varying degrees at several sites in a given polypeptide. Also, a given polypeptide may contain many types of modifications. Polypeptides may be branched, for example, as a result of ubiquitination, and they may be cyclic, with or without branching. Cyclic, branched, and branched cyclic polypeptides may result from posttranslational natural processes or may be made by synthetic methods. Modifications include acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphatidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cysteine,

formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, pegylation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. (See, for instance, PROTEINS - STRUCTURE AND MOLECULAR PROPERTIES, 2nd Ed., T. E. Creighton, W. H. Freeman and Company, New York (1993); POSTTRANSLATIONAL COVALENT MODIFICATION OF PROTEINS, B. C. Johnson, Ed., Academic Press, New York, pgs. 1-12 (1983); Seifter et al., *Meih Enzymol* 182:626-646 (1990); Rattan et al., *Ann NY Acad Sci* 663:48-62 (1992).)

"SEQ ID NO:X" refers to a polynucleotide sequence while "SEQ ID NO:Y" refers to a polypeptide sequence, both sequences identified by an integer specified in Table 1.

"A polypeptide having biological activity" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the present invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. In the case where dose dependency does exist, it need not be identical to that of the polypeptide, but rather substantially similar to the dose-dependence in a given activity as compared to the polypeptide of the present invention (i.e., the candidate polypeptide will exhibit greater activity or not more than about 25-fold less and, preferably, not more than about tenfold less activity, and most preferably, not more than about three-fold less activity relative to the polypeptide of the present invention.)

## **Polynucleotides and Polypeptides of the Invention**

### **FEATURES OF PROTEIN ENCODED BY GENE NO: 1**

This gene is expressed primarily in testes (tumor and to a lesser extent in fetal brain). Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly of the testes, and defects of the central nervous system such as seizure and neurodegenerative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly cancer of the testes and central nervous system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, brain and other tissue of the nervous system, and blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of testicular cancer and treatment of central nervous system disorders since this gene is primarily expressed in the testes tumor and developing brain.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 2

This gene is expressed primarily in cancer tissues, such as breast cancer and Wilms' tumor, and to a lesser extent in fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and/or tumors, particularly, those found in the breast, and developmental abnormalities or disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the glandular tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and fetal tissue and, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 314 as residues: Pro-11 to Thr-18, Leu-43 to Pro-50, Gly-64 to Leu-72, and Leu-81 to Lys-86.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of cancers and/or tumors, particularly, those found in the breast since expression is mainly in cancer/tumor tissues. May serve as therapeutic proteins for proliferation/differentiation of fetal tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 3

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in spleen, chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or leukemias, diseases of the immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders or leukemias, diseases of the immune system since expression is in tissues related to immune function.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 4

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood disorders or lymphocytic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 5

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 317 as residues: Pro-13 to Lys-21.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders since expression is in tissues related to immune function.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 6

This gene is expressed primarily in CD34 depleted buffy coat.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: blood or immune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level

in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 318 as residues: Lys-31 to Lys-39.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood diseases since it is expressed in tissues related to immune function.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 7

This gene is expressed primarily in CD34 depleted buffy coat and to a lesser extent in pineal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system and brain associated diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and pineal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/diagnosis of blood disorders, immune diseases or brain associated diseases (specifically of the pineal gland) since expression is in tissues related to immune function.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 8

The translation product of this gene shares sequence homology with an organic cation transporter which is thought to be important in organic cation uptake in the kidney and liver. (See Accession No. 2343059.) Preferred polypeptide fragments comprise the amino acid sequence ITAIQMICLVNXXEL YPTFVRNXGVNVCSSLCDIGIITP FIVFRLREVWQALPLILFAVLGLLAAGVTLLLPETKGVLPETMKDAENLGRKAKPKENTLYLK VQTSEPSQT (SEQ ID NO: 615) or TMKDAENLGRKAKPKENT (SEQ ID NO: 616) as well as N-terminal and C-terminal deletions of these fragments. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic and renal diseases where drug elimination/cation exchange (organic cation uptake) in the liver and kidney are problematic. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 320 as residues: Asn-64 to Asn-74, and Gln-81 to Gly-87.

The tissue distribution and homology to organic cation transporter indicate that polynucleotides and polypeptides corresponding to this gene are useful as a polyspecific transporter that is important for drug elimination in the liver (and possibly kidney) since expression is found in the liver.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 9

This gene is expressed primarily in eosinophil induced with IL-5 and to a lesser extent in fetal liver and spleen. This gene also maps to chromosome 15, and therefore can be used in linkage analysis as a marker for chromosome 15.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases of the immune system, particularly allergies or asthma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the

standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosis of diseases involving eosinophil reactions since expression seems to be concentrated in eosinophils and other tissues involved in immunity such as the liver and spleen.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 10

This gene is expressed primarily in tissues of hematopoietic lineage and to a lesser extent in Hodgkins lymphoma. Any frame shifts in this sequence can easily be clarified using known molecular biology techniques.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and immune deficiency or dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, lymphoid and reticuloendothelial tissues, and cancerous tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment/ diagnosis for lymphomas or immune dysfunction or as a therapeutic protein useful in immune modulation based on expression in anergic T-cells and lymphomas.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 11

This gene is expressed primarily in neutrophils and to a lesser extent in activated lymphoid cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the cell type present in a biological sample and for diagnosis of diseases and conditions: inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders

of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 323 as residues: Glu-40 to Lys-46.

5

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for modulation of an immune reaction or as a growth factor for the differentiation or proliferation of neutrophils for the treatment of neutropenia.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 12

This gene is expressed primarily in brain and to a lesser extent in activated T-cells. It is likely that the open reading frame containing the predicted signal peptide continues in the 5' direction. Preferred polypeptide fragments comprise the amino acid sequence PRVNSPEDGLSLTGDSCKL (SEQ ID NO:617).

15

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disorders including ischemic shock, alzheimers and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain, and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 324 as residues: Ser-5 to Glu-14, Ile-21 to Pro-35, Ser-65 to Asp-81, Cys-89 to Val-96, Lys-136 to Ser-145, Ile-152 to Met-169, and Arg-189 to Lys-196.

30

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnostic/treatment for cancers of the given tissue or in the treatment of neurological disorders of the CNS.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 13

This gene was also recently cloned by other groups, naming this calcium-activated potassium channel gene, hKCa4. (See Accession No. AF033021, see also, Accession No. 2584866.) This gene is mapped to human chromosome 19q13.2. A second signal sequence likely exists upstream from the predicted signal sequence as described in Table 1. Preferred polypeptide fragments comprise: QADDLQATVAALCYLRGGGPWAG

5

SWLSPKTPGAMGGDLVLGLGALRRKRLL (SEQ NO: 618); or EQEKSLAGWALVLAAXGIGL

MVLHAEMLWFGCSAVNATGHLSDTLWLIPITFLTIGYGDVVVPGTMWQKVCVLCCTGVMGVCC

TALLVAVVARKLEFNKAEKHVHFMMDIQYTKEMKESAAARVLQEAWMFYKHTRRKESHAAR

10

XHQKXLLAAINAFRQVRLKRLREQVNSMVDISKMHMILYDLQNLSSSHRALEKQIDTL

KLDALTELLSTALGPRQLPEPSQSK (SEQ ID NO: 619), as well as N-terminal and C-

terminal deletions. Also preferred are polynucleotide fragments encoding these

polypeptide fragments.

This gene is expressed primarily in breast lymph node and T-cells, and to a

lesser extent in placenta.

15

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: hematologic and

immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides

are useful in providing immunological probes for differential identification of the

tissue(s) or cell type(s). For a number of disorders of the above tissues or cells,

particularly of the immune system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid

tissue, blood cells and placenta, and cancerous and wounded tissues) or bodily fluids

(e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell

sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an

individual not having the disorder. Preferred epitopes include those comprising a

sequence shown in SEQ ID NO. 325 as residues: Arg-13 to Lys-23.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the treatment/diagnosis of hematologic and

diseases involving immune modulation based on distribution in the lymph node and T-

cells.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 14

This gene was recently cloned by another group, calling it PAPS synthase.

(See Accession No. e1204135.) Preferred polypeptide fragments comprise the amino acid sequence YQAHFVSQKRCQVVGTRGFRGCTVWL TGLSGAGK (SEQ ID NO: 620).

Also preferred are the polynucleotide fragments encoding this polypeptide fragment.

It has been discovered that this gene is expressed primarily in benign prostate hyperplasia, Human Umbilical Vein Endothelial Cells and to a lesser extent in smooth muscle and Human endometrial stromal cells-treated with estradiol.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammation, ischemia, and restenosis, based on endothelial cell and smooth muscle cell expression, and prostate diseases such as benign prostate hyperplasia or prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s).

For a number of disorders of the above tissues or cells, particularly of the prostate or vessels of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, endothelial cells, smooth muscle, and endometrium, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 326 as residues: Arg-21 to Asp-26, Lys-35 to Lys-44, Glu-49 to Asn-58.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating/diagnosing diseases or conditions where the endothelial cell lining of the veins and arteries of underlying smooth muscle are involved.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 15

This gene is expressed primarily in human 6 week embryo and to a lesser extent in placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental anomalies or fetal deficiencies. Similarly, polypeptides and antibodies directed to these

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly developmental in nature, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 327 as residues Lys-50 to Glu-57.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection of developmental abnormalities.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 16

This gene is expressed primarily in kidney and amygdala and to a lesser extent in fetal tissues. This gene is mapped to chromosome 14, and therefore is useful in linkage analysis as a marker for chromosome 14.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) present in a biological sample and for diagnosis of diseases and conditions: kidney diseases, neurological disorders and developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s). For a number of disorders of the above tissues, particularly of the renal system or developing fetal tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, amygdala, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of conditions affecting the brain, kidneys and fetal development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 17

This gene is expressed primarily in ovarian cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: solid tumors similar to ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 329 as residues Ser-51 to Val-56.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment of solid tumors of the reproductive system such as ovarian cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 18

This gene is expressed primarily in brain medulloblastoma. Preferred polypeptide fragments comprise the amino acid sequence: IRHEQHFNFSLEMHSGSLLFLPLQL ILLPVCAHLHEELNC (SEQ ID NO: 643) and SPTSEEKGHLLQLQAERHPWVAGALVGYSGLTLTCSGPTEKPAKGYFLKLLQEMHIRAN (SEQ ID NO: 644), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the CNS or. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating medulloblastoma or similar tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 19

This gene is expressed primarily in adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose tissues expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating obesity by regulating the function and number of adipocytes

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 20

This gene is expressed primarily in B cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of the immune system with an emphasis on B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the tumors of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,



the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of B cell derived tumors based on its expression in b cell lymphomas

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 21

This gene is expressed primarily in immune cells and to a lesser extent in fetal tissues

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cells of the immune system, and fetal tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO:333 as residues Asp-10 to Pro-19, Ser-74 to Tyr-79, Glu-95 to Lys-110.

The tissue distribution indicates that polynucleotides and polypeptides

25 corresponding to this gene are useful for treatment of diseases involving alterations in T cell activity.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 22

It has been discovered that this gene is expressed primarily in ovarian tumor.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors particularly of the ovary. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of tumors of the reproductive organs. expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovarian

and other reproductive tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO. 334 as residues: Leu-22 to Gln-27.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovarian tumors as it has only been identified in ovarian tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 23

It has been discovered that this gene is expressed primarily in fetal tissues and to a lesser extent in osteoclastoma cell line

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: osteoporosis or arthritis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of conditions of abnormal bone remodeling due to enhanced activity of osteoclasts. This may be useful as a specific marker for malignancies derived from osteoclasts or their precursors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 24

The translation product of this gene shares sequence homology with a periplasmic ribonuclease which is thought to be important in degrading extracellular polynucleotides

35 It has been discovered that this gene is expressed primarily in serum treated smooth muscle cells

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular disease such as stenosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculature expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. Preferred epitopes include those comprising a sequence shown in SEQ ID NO: 336 as residues: Gln-30 to Lys-36, and Pro-41 to Arg-48.

The tissue distribution and homology to ribonucleases indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of pathological conditions of smooth muscle associated with bacterial or viral infiltration

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 25

This gene is expressed primarily in Early Stage Human Brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain development and related diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain development and related diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases affecting human brain development and related diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 26

It has been discovered that this gene is expressed primarily in human brain tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases and other diseases related to brain diseases, which may be caused by brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human brain diseases and other diseases related.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 27

It has been discovered that this gene is expressed primarily in Anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases, inflammatory diseases and diseases related to T lymph cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune diseases, inflammatory diseases and diseases related to T lymph cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene

expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for immune diseases, and inflammatory diseases and diseases related to T lymph cells.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 28

The translation product of this gene shares sequence homology with *Shigella flexneri* positive transcriptional regulator CtrR (ctrR) gene which is thought to be important in regulation of gene expression.

10

This gene is expressed primarily in human synovial sarcoma and normal human brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: human brain diseases particularly sarcomas of the synovium. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the human brain and synovium and other related human brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., synovial tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of human synovial sarcoma and other related human brain diseases.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 29

This gene is expressed in bone marrow, infant brain, fetal liver and spleen, prostate and to a lesser extent in pineal gland, adipose tissue, kidney, adrenal gland, umbilical vein endothelial cells, and T cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases related to bone marrow or

35

hematoplasic tissues, prostate, kidney, adrenal gland, and cardiovascular tissue or organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases related to hematoplasic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, hematopoietic cells, pineal gland, adipose tissue, kidney, adrenal gland, endothelial cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5

10

The tissue distribution and homology to the gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases related to hematoplasic tissues, immune system, prostate, kidney, adrenal gland, and cardiovascular tissue or organs.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 30

This gene is expressed primarily in meningea and to a lesser extent in breast and adult brain.

20

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases of the meningea and related brain diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above

25

tissues or cells, particularly of the meningea and related brain diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., meningea, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the

30

35

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of diseases of the meningea and related brain diseases.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 31

This gene is expressed in meningea, fetal spleen, osteoblast and to a lesser extent in activated T-cells, endometrial stromal cells, fetal lung, HL-60, thymus, testis and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: meningeal disease, osteoporosis, immune diseases, and hematoplastic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the

15 above tissues or cells, particularly of the meningeal diseases, osteoporosis, immune diseases, and hematoplastic diseases, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, endometrium, lung, thymus, testis, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of meningeal, osteoporosis, immune diseases, hematoplastic diseases, testis diseases and lung diseases.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 32

This gene is expressed primarily in human thymus and to a much lesser extent in infant brain, T-cells, smooth muscle, endothelial cells, bone marrow, human ovarian tumor and keratinocytes testes, osteoclastoma, breast, and tonsils.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Diseases involving the thymus, particularly thymic cancer and diseases involving T-cell maturation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a

35

number of disorders of the above tissues or cells, particularly of the thymus, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, brain, and other tissue of the nervous system, blood cells, bone marrow, ovaries, and testes, and other reproductive tissue, mammary tissue, tonsils, melanocytes and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5

The tissue distribution and homology to gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases of the thymus particularly thymic cancer and diseases involving T-cell maturation.

10

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 33

This gene is expressed primarily in human tonsils, and placenta, and to a lesser extent in adipocytes, melanocyte, and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: inflammatory diseases, immune diseases, and obesity. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the inflammatory diseases, immune diseases, and obesity, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., tonsils, placenta, adipocytes, melanocytes, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

25

30

The tissue distribution and homology to this gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of diseases such as inflammation, immune diseases, and obesity.

35

**FEATURES OF PROTEIN ENCODED BY GENE NO: 34**

This gene is expressed in activated T cells, and to a lesser extent in pituitary, testis, and breast lymph node.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases relating to T cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the disorders of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary, testes and other reproductive tissue, mammary tissue, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of immune disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 35**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain, and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 36**

This gene is expressed primarily in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as

- 5 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the diseases relating to neurological disorders, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of neurological disorders.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 37**

This gene is expressed primarily in human ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as

- reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the ovarian disorders such as those involving germ cells, ovarian follicles, stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of ovaropathy.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 38**

This gene is expressed primarily in lymph node breast cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the breast cancer, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for used as a diagnostic marker for breast cancer.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 39**

This gene is expressed primarily in brain and to a lesser extent in other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal disorders such as trauma, brain degeneration, and brain tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of neuronal disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 40**

This gene is expressed in early stage human embryo, adrenal gland tumor, and immune tissues such as fetal liver, fetal spleen, T-cell, and myeloid progenitor cell line and to a lesser extent in ovary, colon cancer, and a few other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis including adrenal gland tumor, colon cancer and various other tumors, developmental and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the cancer tissues, early stage human tissues, and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, blood cells, bone marrow, ovary and other reproductive tissue, and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and therapeutic treatment of immune and developmental disorders, and tumorigenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 41**

This gene is expressed primarily in fetal lung, endothelial cells, liver, thymus and a few other immune tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders such as immune deficiency and autoimmune diseases, pulmonary diseases, liver diseases, and tumor metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal lung, liver, endothelial cells, and immune tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain

5 tissues and cell types (e.g., lung, endothelial cells, liver, thymus, and other tissue of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of immune disorders and pulmonary and hepatic diseases. Its promoter may also be used for immune system and lung-specific gene therapies. The expression of this gene in endothelial cells indicates that it may also involve in angiogenesis which therefore may play role in tumor metastasis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 42

15 This gene is expressed primarily in liver, thyroid, parathyroid and to a lesser extent in fetal lung, stomach and early embryos.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic regulation, obesity, hepatic failure, hepatocellular tumors or thyroiditis and thyroid tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive/endocrine system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, thyroid, parathyroid, lung, stomach, and embryonic tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and the extracellular locations indicates that polynucleotides and polypeptides corresponding to this gene are useful for the detection and treatment of digestive/endocrine disorders, including metabolic regulation, hepatic failure, malabsorption, gastritis and neoplasms.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 43

This gene is expressed primarily in Schizophrenic adult brain, pituitary, front cortex, hypothalamus and to a lesser extent in retina, adipose and stomach cancer and placenta.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal tissue, adipose, stomach, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in treatment/detection of disorders in the nerve system, including schizophrenia, neurodegeneration, and neoplasia. Additionally, a secreted protein in brain may serve as an endocrine.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 44

25 The translation product of this gene shares sequence homology with GTP binding proteins which are thought to be important in signal transduction and protein transport.

This gene is expressed primarily in umbilical vein and microvascular endothelial cells, GM-CSF treated macrophage, anergic T cells, osteoblast, osteoclast, CD34+ cells and to a lesser extent in gall bladder.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: bone formation and growth, osteonecrosis, osteoporosis, angiogenesis and/or hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoiesis systems, expression of this gene at significantly higher or lower levels

may be routinely detected in certain tissues and cell types (e.g., endothelial cells, blood cells, bone, and gall bladder, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment/detection of bone formation and growth, osteonecrosis, osteoporosis, and/or hematopoiesis because its involvement in the growth signaling or angiogenesis.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 45

The translation product of this gene shares sequence homology with signal sequence receptor gamma subunit which is thought to be important in protein translocation on endoplasmic reticulum.

This gene is expressed primarily in adrenal gland, salivary gland, prostate, and to a lesser extent in endothelial cells and smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: protein secretion. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the secretory organs, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, salivary gland, prostate, endothelial cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to SSR gamma subunit indicate that polynucleotides and polypeptides corresponding to this gene are useful for endocrine disorders, prostate cancer, xerostomia or sialorrhea.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 46

This gene is expressed primarily in osteoclastoma cells and to a lesser extent in melanocyte, amygdala, brain, and stomach.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ossification, osteoporosis, fracture, osteonecrosis, osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., melanocytes, amygdala, brain and other tissue of the nervous system, and stomach, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in intervention of ossification, osteoporosis, fracture, osteonecrosis and osteosarcoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 48

The translation product of this gene shares sequence homology with proline rich proteins which is thought to be important in protein-protein interaction.

This gene is expressed primarily in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological and psychological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nerve system and endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful in intervention



and detection of neurological diseases, including trauma, neoplasia, degenerative or metabolic conditions in the central nerve system. Additionally, the gene product may be secreted by the brain as an endocrine.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 49

The translation product of this gene shares sequence homology with the AOCB gene from *Aspergillus nidulans* which is important in asexual development.

This gene is expressed primarily in infant brain and to a lesser extent in the

developing embryo, trachea tumors, B-cell lymphoma and synovial sarcoma.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative diseases, leukemia and sarcoma's. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, blood cells, trachea, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution in infant brain and sarcoma's and homology to a gene involved in a key step of eukaryotic development (fungal spore formation) indicates that the protein product of this clone could play a role in neurological diseases such as schizophrenia, particularly in infants. The existence of the gene in a B-cell lymphoma indicates the gene may be used in the treatment and detection of leukemia.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 50

This gene is expressed primarily in fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary disorders including lung cancer. Similarly, polypeptides and antibodies directed to these

35 polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the pulmonary system, expression of this gene at significantly higher or

lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution of this gene only in fetal lung indicates that it plays a key role in development of the pulmonary system. This would suggest that misregulation of the expression of this protein product in the adult could lead to lymphoma or sarcoma formation, particularly in the lung. It may also be involved in predisposition to certain pulmonary defects such as pulmonary edema and embolism, bronchitis and cystic fibrosis.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 51

This gene is expressed primarily in hematopoietic cell types and fetal cells and to a lesser extent in all tissue types.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in the immune system and hematopoiesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution of this gene predominantly in hematopoietic cells and in the developing embryo indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of lymphomas and disease states affecting the immune system or hematopoiesis disorders such as leukemia, AIDS, arthritis and asthma.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 52

This gene is expressed primarily in prostate and to a lesser extent in fetal spleen, fetal liver, infant brain and T cell leukemias.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate disorders, prostate cancer, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, and/or prostate gland expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., thymus, spleen, liver, brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution of this gene in prostate indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection or treatment of prostate disorders or prostate cancer. Its distribution in fetal liver and fetal spleen indicates it may play a role in the immune system and its misregulation could lead to immune disorders such as leukemia, arthritis and asthma.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 53

The translation product of this gene shares sequence homology with dynein. This gene is expressed primarily in brain.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuro-degenerative diseases of the brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly neuro-degenerative diseases expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35

The predominant tissue distribution in the brain and homology to dynein, a microtubule motor protein involved in the positioning of cellular organelles and molecules indicates that polynucleotides and polypeptides corresponding to this gene are useful for detection/treatment of neurodegenerative diseases, such as Alzheimers, 5 Huntingtons, Parkinsons diseases and shizophrenia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 54

The translation product of this gene shares sequence homology with ubiquitin-conjugation protein, an enzyme which is thought to be important in the processing of the Huntingtons Disease causing gene.

10 This gene is expressed primarily in brain and to a lesser extent in activated macrophages.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurodegenerative disease states including Huntington's disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of brain tissues. For a number of disorders of the above tissues or cells, particularly of the neurological systems expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The predominant tissue distribution of this gene in the brain and its homology to a Huntington interacting protein indicates that polynucleotides and polypeptides corresponding to this gene are useful for the regulation of the expression of the Huntington disease gene and other neurodegenerative diseases including spinocerebellar ataxia types I and III, dentatorubropallidoluysian and spinal bulbar muscular atrophy. In addition, the existence of elevated levels of free ubiquitin pools in Alzheimer's disease, Parkinson's disease and amyotrophic lateral sclerosis indicates that the ubiquitin pathway of protein degradation plays a role in these disease states. Thus, considering the gene described here is homologous to a ubiquitin-conjugation 35 protein it may play a general role in neurodegenerative conditions.

# **FEATURES OF PROTEIN ENCODED BY GENE NO: 56**

This gene is expressed primarily in T-cells (anergic T-cells, resting T-Cells, apoptotic T-cells) and lymph node (breast), as well as brain (hypothalamus, hippocampus, pituitary, infant brain, early-stage brain).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune (e.g. immunodeficiencies, autoimmuneities, inflammation, leukemias & lymphomas) and neurological (e.g. Alzheimer's disease, dementia, schizophrenia) disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous, hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, lymphoid tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention or detection of pathologies associated with the hematopoietic and immune systems, such as anemias (leukemias). In addition, the expression in brain (including fetal) might suggest a role in developmental brain defects, neuro-degenerative diseases or behavioral abnormalities (e.g. schizophrenia, Alzheimer's, dementia, depression, etc.).

# **FEATURES OF PROTEIN ENCODED BY GENE NO: 57**

This gene is expressed primarily in lung, and to a lesser extent in a variety of other hematological cell types (e.g. Raji cells, bone marrow cell line, activated monocytes).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: pulmonary and/or hematological dysfunction. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vasculo-pulmonary and hematopoietic systems, expression of this

gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the intervention and detection of pathologies associated with the vasculo-pulmonary system. In addition the expression of this gene in a variety of leukocytic cell types and a bone marrow cell line might suggest a role in hematopoietic and immune system disorders, such as leukemias & lymphomas, inflammation, immunodeficiencies and autoimmuneities.

# **FEATURES OF PROTEIN ENCODED BY GENE NO: 58**

The translation product of this gene shares sequence homology with adenylate kinase isozyme 3 (gII 63528 GTP:AMP phosphotransferase (EC 2.7.4.10) [Bos taurus]), which is thought to be important in catalyzing the phosphorylation of AMP to ADP in the presence of ATP or inorganic triphosphate.

This gene is expressed primarily in fetal liver, heart and placenta, and to a lesser extent in many other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatic, cardiovascular or reproductive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic, cardiovascular and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, heart, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of conditions related to hepatic function and pathogenesis, in particular, those dealing with liver development and the differentiation of hepatocyte progenitor cells.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 59**

This gene is expressed primarily in CD34 positive cells (Cord Blood).

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic differentiation and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in the detection and treatment of conditions associated with CD34-positive cells, and therefore as a marker for cell differentiation in hematopoiesis, as well as immunological disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 60**

The translation product of the predicted open reading frame of this contig has sequence identity to the murine gene designated Insulin-Like Growth Factor-Binding Protein (IGFBP)-1 as described by Lee and colleagues (Hepatology 19 (3), 656-665 (1994)).

This gene is expressed exclusively in hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of hemangiopericytoma and other pericyte or endothelial cell proliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and immune systems, expression of this gene at significantly higher or lower levels may routinely be detected in certain tissues and cell types (e.g., pericyte or endothelial cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or

another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

Polynucleotides and polypeptides corresponding to this gene are useful as cell growth regulators since IGFBP-1-like molecules function as modulators of insulin-like growth factor activity. In addition, since IGFBP-1 is expressed at high levels following hepatectomy and during fetal liver development, polynucleotides of the present invention may also be used for the diagnosis of developmental disorders. Further, polypeptides of the present invention may be used therapeutically to treat developmental liver disorders as well as to regulate hepatocyte and supporting cell growth following hepatectomy or to treat liver disorders.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hemangiopericytoma and liver disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 61**

This gene is expressed primarily in schizophrenic frontal cortex.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: nervous system and cognitive disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the frontal cortex and CNS expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of frontal cortex, neuro-degenerative and CNS disorders

**FEATURES OF PROTEIN ENCODED BY GENE NO: 62**

This gene is expressed primarily in human adrenal gland tumor, and to a lesser extent in human kidney, medulla and adult pulmonary tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic, endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and nervous system disorders and neoplasia, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adrenal gland, kidney, brain and other tissue of the nervous system, pulmonary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, treatment and diagnosis of neurological and endocrine disorders including neoplasia.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 63

This gene is expressed primarily in human adipocytes, and to a lesser extent in spleen, 12-week old human, and testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune, metabolic and growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipocytes, spleen, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of immune, developmental and metabolic disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 64

One translated product of this clone is homologous to the mouse zinc finger protein PZF. (See Accession No. 453376; see also Gene 152 (2), 233-238 (1995).) Preferred

polypeptide fragments correspond to the highly conserved domains shared between mouse and man. For example, preferred polypeptide fragments comprise the amino acid

sequence: LQCEICGFTCRQKASLNWHMKKHDDADSFYQFSCNCGKKFEEKKDSVVVAKAKSH

PEV (SEQ ID NO: 621); ITSTDILGTNPESLTQPSD (SEQ ID NO: 622); NSTSGECLLLEAGQM

SKSY (SEQ ID NO: 623); CSQTERVSLMADGKIFVGGSSGGTGLVMNSDILGATTEVLIEDSD

SAGP (SEQ ID NO: 624); IQYVRCEMCGGTVLAHPRYLQHHLKKYVCPHPSCGRLF

RLQQLLRHAKHHT (SEQ ID NO: 625); DQRDYCEYCARAFKSSHNLAVHRMIHTGEK (SEQ

ID NO: 626); RSSRTSVSRHRDTESTRSSKSTGSLQICKSEPTDQLDY (SEQ ID NO: 627);

PFKDDPRDETYPHLERETPKRRKSG (SEQ ID NO: 630); QYVRCEMCGGTVLAHPRYLQ

HHIKYQHLLKKYVCPHPSCGRLFRLQQLLRHAKHHTD (SEQ ID NO: 629); or residues

151-182 of QRDYCEYCARAFKSSHNLAVHRMIHTGEKHY (SEQ ID NO: 628). Also preferred are polynucleotide fragments encoding these polypeptide fragments.

This gene is expressed primarily in Rhabdomyosarcoma, melanocyte and colon cancer tissue and to a lesser extent in smooth muscle, pancreatic tumor, and apoptotic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and hemopoietic, expression of this gene at significantly higher or lower

levels may be routinely detected in certain tissues and cell types (e.g., striated muscle, melanocytes, colon, smooth muscle, pancreas, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal

fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis and treatment of cancer and hemopoietic disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 65**

This gene is expressed primarily in human adipose and salivary gland tissue and to a lesser extent in human bone marrow and fetal kidney.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: metabolic and immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the metabolic and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose, salivary gland, bone marrow, and kidney, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for study, diagnosis of metabolic and immune disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 66**

This translated product of this gene was recently identified as oxytocinase splice variant 1. (See Accession Nos. 2209276 and d1010078.) Preferred polypeptide fragments

comprise the amino acid sequence: EMFDSLVSFKGSSLLMLKTYLSEDFVQHA VVLYLHN HSYASIQSDDLWDSFNEVTNQTLDVKRMKMTWTLQKGFPLVTYVQKKGKELFQQRFFLNMK PEIQPSDTRYM (SEQ ID NO: 631). Also preferred are polynucleotide fragments encoding this polypeptide fragment.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 67**

This gene is expressed primarily in hemopoietic cells, particularly apoptotic T-cells, and to lesser extent in primary dendritic cells and adipose tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of apoptotic T-cells, primary dendritic cells, and adipose tissue present in a biological sample and for diagnosis of diseases and conditions: hemopoietic diseases including cancer and general immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the oral and intestinal mucosa as well as hemopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of diseases of the immune system, including cancer, hemopoietic and infectious diseases.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 68**

This gene is expressed primarily in kidney cortex and to a lesser extent in infant brain, heart, uterus, and blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of kidney tissue present in a biological sample and for diagnosis of diseases and conditions: soft tissue cancer, inflammation, kidney fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and endocrines systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, brain, and other nervous tissue, heart, uterus, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and fibroses.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 69**

The translation product of this gene shares strong sequence homology with vertebrate and invertebrate protein tyrosine phosphatases.

This gene is expressed primarily in endometrial tumors, melanocytes, myeloid progenitors and to a lesser extent in infant brain, adipocytes, and several hematopoietic stem cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of transformed hematopoietic and epithelial cells present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of skin and endometrium, leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, melanocytes, bone marrow, adipocytes, hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and sequence similarity with tyrosine phosphatases

indicate that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of cancer and hematopoietic disorders.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 70

This gene is expressed primarily in osteoclastoma, breast, and infant brain and to a lesser extent in various fetal and transformed bone, ovarian, and neuronal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: degenerative conditions of the brain and skeleton. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and skeletal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, mammary tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of degenerative, neurological and skeletal disorders.

## 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 71

This gene was originally cloned from tumor cell lines. Recently another group has also cloned this gene, calling it the human malignant melanoma metastasis-suppressor (KISS-1) gene. (See Accession No. U43527.) Preferred polypeptide fragments comprise the amino acid sequence: LEKVASVGNRPQTQQLESIGLLA (SEQ ID NO: 632); VHREASCYCQAEPSGDL (SEQ ID NO: 633); RPALRQAGGOTREPRQKRWAGL (SEQ ID NO: 634); and AVNFRPQRSQM (SEQ ID NO: 635). Any frame shifts can easily be resolved using known molecular biology techniques.

This gene is expressed primarily in many types of carcinomas and to a lesser extent in many normal organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanomas, and other hyperproliferative disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of transformed organ tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder. As a tumor suppressor gene, increase amounts of the polypeptide can be used to treat patients having a particular cancer.

The tissue distribution indicates that this gene and the translated product is useful for diagnosing and study of cancer.

## 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 72

This gene is expressed primarily in striatum and to a lesser extent in adipocytes and hemangiopericytoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of striatal cells present in a biological sample and for diagnosis of diseases and conditions: neurological, fat and lysosomal storage

diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., striatal tissue, adipocytes, and vascular tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of neurodegenerative and growth disorders.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 73

This gene is expressed primarily in bone marrow stromal cells and to a lesser extent in smooth muscle, testes, endothelium, and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of bone marrow present in a biological sample and for diagnosis of diseases and conditions: connective tissue and hematopoietic diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, stromal cells, smooth muscle, testes and other reproductive tissue, endothelium, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of connective tissue and blood diseases.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 74

This gene is expressed primarily in brain, fetal liver and lung and to a lesser extent in retina, spinal chord, activated T-cells and endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of brain and regenerating liver present in a biological sample and for diagnosis of diseases and conditions: CNS and spinal chord injuries, immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, pulmonary tissue, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study and treatment of hematopoietic and neurological conditions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 75

The translation product of this gene shares sequence homology with GTP binding proteins (intracellular).

This gene is expressed primarily in bone marrow, brain, and melanocytes and to a lesser extent in various endocrine and hematopoietic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hematopoietic and nervous system conditions. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, melanocytes, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder.



relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to nucleotide binding factors indicate that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of brain degenerative, skin and blood diseases.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 76

This gene is expressed primarily in activated T-cells and to a lesser extent in retina, brain, and fetal bone.

10

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of activated T-cells and developing brain present in a biological sample and for diagnosis of diseases and conditions: immune deficiencies and skeletal and neuronal growth disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and skeletomuscular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, brain and other tissue of the nervous system, retinal tissue, and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis, study and treatment of cancer, urogenital, and brain degenerative diseases.

25

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 77

This gene is expressed primarily in fetal liver, activated monocytes, osteoblasts and to a lesser extent in synovial, brain, and lymphoid tissues.

30

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of myeloid and lymphoid present in a biological sample and for diagnosis of diseases and conditions: inflammation, immune deficiencies, cancer. Similarly, polypeptides and antibodies directed to these

35

polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and skeleton, expression of this gene at significantly

higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, blood cells, bone, synovial tissue, brain and other tissue of the nervous system, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for study, diagnosis, and treatment of lymphoid and mesenchymal cancers and nervous system diseases.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 78

The translation product of this gene shares sequence homology with polymerase polypeptide precursor which is thought to be important in DNA repair and replication. This gene is expressed primarily in infant brain and to a lesser extent in tumors and tumor cell lines.

15

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, especially of the neural system and developing organs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

30

The tissue distribution and homology to polymerase polypeptide precursor indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers especially of the neural system and developing organs

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 79

This gene is expressed primarily in muscle and endothelial cells and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: vascular diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., muscle, endothelial cells, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for treatment and diagnosis of disorders of the vascular and neural system including cardiovascular and endothelial.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 80

This gene is expressed primarily in placenta and to a lesser extent in fetal liver

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: developmental disorders and disorder of the haemopoietic system, fetal liver and placenta. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developmental disorders and disorder of the haemopoietic system, fetal liver and placenta, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of developmental disorders and disorders of the haemopoietic system, fetal liver and placenta.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 81

This gene is expressed primarily in bone marrow, placenta and tissues and organs of the hematopoietic system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the bone and haemopoietic system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, bone and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone marrow, placenta, and hematopoietic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis and treatment of disorders of the immune, bone and hematopoietic system

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 82

The translation product of this gene shares sequence homology with secretory carrier membrane protein which is thought to be important in protein transport and export. Any frame shifts in coding sequence can be easily resolved using standard molecular biology techniques. Another group recently cloned this gene, calling it SCAMP. (See Accession No. 2232243.)

This gene is expressed primarily in prostate, breast and spleen, and to a lesser extent in several other tissues and organs.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorders of the breast prostate and spleen. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly disorders of the breast prostate and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell

types (e.g., prostate, mammary tissue, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secretory carrier membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of disorders of the breast, prostate and spleen.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 83

This gene is expressed primarily in developing organs and tissue like placenta and infant brain and to a lesser extent in developed organs and tissue like cerebellum and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as

15 reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, heart, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of diseases of the neural system including neurological disorders and cancer.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 84

The translation product of this gene shares sequence homology with ATPase 6 in *Trypanosoma brucei* which is thought to be important in metabolism.

This gene is expressed primarily in tumor and fetal tissues and to a lesser extent in melanocytes, kidney cortex, monocytes and ovary.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: metabolism disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, melanocytes, kidney, blood cells, ovary and other tissue of the reproductive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ATPase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of metabolism disorders, especially in fetal and tumor tissue growth.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 85

The translation product of this gene shares sequence homology with the immunoglobulin superfamily of proteins which are known to be important in immune response and immunity.

This gene is expressed primarily in stromal cells, colon cancer, lung, amygdala, melanocyte and to a lesser extent in a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the stromal cells, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells, colon, lung, amygdala, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of immune system disorders.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 86**

The translation product of this gene shares sequence homology with transcription initiation factor eIF-4 gamma which is thought to be important in gene transcription.

This gene is expressed primarily in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumorigenesis.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in tumor tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to transcription initiation factor eIF-4 gamma indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene regulation in tumorigenesis.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 87**

The translation product of this gene shares sequence homology at low level in prolines with secreted basic proline-rich peptide II-2 which is thought to be important in protein structure or inhibiting hydroxyapatite formation in vitro.

This gene is expressed primarily in endometrial tumor and fetal lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: endometrial tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular/skeletal and reproductive systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample

taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to secreted basic proline-rich peptide II-2 indicate that polynucleotides and polypeptides corresponding to this gene are useful for inhibiting hydroxyapatite formation or establishing cell/tissue structure.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 88**

This gene is expressed primarily in: amniotic cells induced with TNF in culture; and to a lesser extent in colon tissue from a patient with Crohn's Disease; parathyroid tumor; activated T-cells; cells of the human Caco-2 cell line; adenocarcinoma; colon; corpus colosum; fetal kidney; pancreas tumor; fetal brain; early stage brain, and anergic T-cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system; e.g., tumors, expression of this gene at significantly higher or lower levels may be routinely detected in certain (e.g., amniotic cells, colon, kidney, pancreas, parathyroid, brain and other tissue of the nervous system, blood cells, hematopoietic cells, liver, spleen, bone, testes and other reproductive tissue, brain and other tissue of the nervous system, and epithelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for modulating tumorigenesis and other immune system conditions such as disorders in immune response.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 89**

This gene is expressed primarily in fetal liver/spleen and hematopoietic cells and to a lesser extent in brain, osteosarcoma, and testis tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

biological sample and for diagnosis of diseases and conditions: leukemia and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, liver, spleen, bone, testes, and other reproductive tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of hematopoietic and immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 90

The translation product of this gene shares weak sequence homology with mouse Gcapi protein which is developmentally regulated in brain.

This gene is expressed primarily in infant and adult brain and fetal liver/spleen and to a lesser extent in smooth muscle, T cells, and a variety of other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune, and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, spleen, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and its homology to Gcapi protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosis of disorders in neuronal, hematopoietic, immune, and endocrine systems.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 91

This gene is expressed primarily in brain and hematopoietic cells and to a lesser extent in tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disorder in nervous, hematopoietic, immune systems and tumorigenesis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, hematopoietic, immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of disorders in the nervous, hematopoietic, and immune systems.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 92

The translation product of this gene shares sequence homology with neuroendocrine-specific protein A which is thought to be important in neurologic systems.

This gene is expressed primarily in brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neural disorders and degeneration disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central or peripheral nervous systems, expression of this gene at

significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., hematopoietic cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to neuroendocrine-specific protein A indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of neural disorders and degeneration disease.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 93

The translation product of this gene shares sequence homology with collagen-like protein and prolin-rich protein which are thought to be important in connective tissue function and tissue structure.

15

This gene is expressed primarily in fetal liver/spleen and brain tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuronal or hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and hematopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to collagen-like protein and proline-rich proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for supporting brain and hematopoietic tissue function and diagnosis and treatment of disorders in these functions.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 94

This gene is expressed primarily in embryonic tissues and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a

35

biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system (e.g., tumors), expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancer.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 95

This gene is expressed primarily in brain tumor, placenta, and melanoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor or melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain or melanocytes, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, placenta, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20

The tissue distribution indicates that the translation product of this gene is useful in the diagnosis and treatment of brain tumors and melanoma.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 96

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

30

The translation product of this gene shares sequence homology with a yeast membrane protein, SUR4, which encodes for APA1 that acts on a glucose-signaling pathway that controls the expression of several genes that are transcriptionally regulated by glucose.

35

This gene is expressed primarily in fetal liver, and to a lesser extent in placenta and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of fetal liver or defects of glucose-regulated ATPase activities in tissues. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune/hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, placenta, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to yeast SUR4 membrane protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of defects of fetal liver or defects of glucose-regulated ATPase activities.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 97

This gene is expressed primarily in fetal liver, brain, and amniotic fluid.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the fetal immune system and adult brain. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal immune system and adult brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., amniotic fluid, serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the fetal immune and hematopoietic systems since fetal liver is

the predominant organ responsible for hematopoiesis in the fetus. In addition, the gene product of this gene is thought to be useful for detecting certain neurological defects of the brain.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 98

The translation product of this gene shares sequence homology with an yolk protein precursor, Vitellogenin which is thought to be important in binding lipids such as phosphatidylcholine.

This gene is expressed primarily in amniotic cells and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in amniotic cells, fetal liver development and the fetal immune system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the [insert system where a related disease state is likely, e.g., immune], expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to vitellogenin indicate that the protein product of this clone is useful for treatment and diagnosis of defects in amniotic cells, fetal liver development and the fetal immune system.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 99

This gene is expressed primarily in placenta, endometrial tumor, osteosarcoma and stromal cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumor of the endometrium or bone, and osteosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the obstetric system (e.g. placenta,

endometrium) and the bones, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endometrium, bone, and stromal cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors and abnormalities of the endometrium, and the bones because of its abundance in the aforementioned tissues.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 100

This gene is expressed primarily in hepatocellular tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hepatocellular tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the liver, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of hepatocellular cancer because of its abundant expression in this tissue.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 101

This gene is expressed primarily in Corpus Colosum, fetal lung and infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the Corpus Colosum or defects of the fetal lung. Similarly, polypeptides and antibodies directed to

these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the Corpus Colosum and brain in general, and fetal lung, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lung, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of the Corpus Colosum and brain in general, and defects of fetal lung.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 102

This gene is expressed primarily in T cells and stromal cells, and to a lesser extent in adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of T cell immunity and stromal cell development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, stromal cells, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of defects of T cell immunity and stromal cell development because of its abundant expression in these tissues.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 103

This gene is expressed primarily in infant brain and placenta.



Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of the brain and nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, especially brain, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for detecting defects of the brain, especially in young children.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 105

This gene is expressed primarily in human osteoclastoma and to a lesser extent in human pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly osteoclastoma and pancreatic tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in transformed tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone and pancreas, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of some types of tumors, particularly pancreatic cancer and osteoclastoma.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 106

This gene is expressed primarily in fetal liver/spleen, and to a lesser extent in activated T-Cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis or treatment of immune disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 107

This gene is expressed primarily in human embryo and to a lesser extent in spleen and chronic lymphocytic leukemia.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: leukemia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune or hemopoietic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, spleen, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the diagnosis and treatment of leukemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 108**

This gene is expressed primarily in placenta, and to a lesser extent in early stage human brain and in lung.

5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: fetal developmental abnormalities. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly in fetal and amniotic tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this is useful for production of growth factor(s) associated with fetal development. Preferred polypeptides comprise the full-length polypeptide shown in the sequence listing, truncated however, at the amino terminus and beginning with QTIE.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 109**

This gene is expressed primarily in fetal spleen, and to a lesser extent in B-Cell lymphoma and T-Cell lymphoma.

25 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for the treatment and diagnosis of human lymphomas.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 110**

5 The translation product of this gene shares sequence homology with sarcoma amplified sequence (SAS), a tetraspan receptor which is thought to be important in malignant fibrous histiocytoma and liposarcoma.

This gene is expressed primarily in human osteoclastoma, and to a lesser extent in pineal gland and infant brain.

10 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: malignant fibrous histiocytoma and liposarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, pineal gland, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to sarcoma amplified sequence (SAS) indicate that the protein product of this clone is useful for treatment of, osteosarcoma, malignant fibrous histiocytoma and liposarcoma and related cancers, particularly sarcomas.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 111**

The translation product of this gene shares sequence homology with 6.8K proteolipid protein, mitochondrial - bovine.

30 This gene is expressed primarily in Wilm's tumor and to a lesser extent in cerebellum and placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Wilm's tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the immune or renal systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to 6.8K proteolipid protein indicate that the protein product of this clone is useful for diagnostic and therapeutics associated with tumors, particularly Wilm's tumor disease.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 112

This gene is expressed primarily in embryonic tissue and to a lesser extent in osteoblasts, endothelial cells, macrophages (GM-CSF treated), and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, bone, endothelial cells, blood cells and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment or diagnosis of immune disorders. Preferred polypeptides encoded by this gene comprise the following amino acid sequence: MITDQLAIFANMLGVSLFLVLYHYVAVNNPKKQE (SEQ ID NO: 636).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 113

This gene is expressed primarily in hepatocellular tumor, and to a lesser extent in fetal liver/spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors, particularly hepatocellular tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of tumors, particularly hepatocellular tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 114

The translation product of this gene exhibits a very high degree of sequence identity with the human Pig8 gene which is thought to be important in p53 mediated apoptosis. The sequence of this gene has since been published by Polyak and colleagues (Nature 389, 300-306 (1997)). In addition, the predicted translation product of this contig exhibits very high sequence homology with a murine gene denoted as EI24 which is also thought to be important in p53 mediated apoptosis.

This gene is expressed primarily in infant brain and activated T-cells and to a lesser extent in bone marrow, fetal liver, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, and tissue damage by radiation and anti-cancer drugs. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, bone marrow, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder,

relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to human Pig8 and murine E124 genes

indicate that polynucleotides and polypeptides corresponding to this gene are useful for preventing apoptosis in patients being treated with anti-oncogenic drugs such as etoposide, hydroperoxycyclophosphamide, and X-irradiation, since this protein product is upregulated in cells undergoing such treatment where p53 was overexpressed. It may also be useful in the treatment of hematopoietic disorders and in boosting numbers of hematopoietic stem cells by interfering with the apoptosis of progenitor cells. The mature polypeptide is predicted to comprise the following amino acid sequence:

10 EEMADSVKTFLDLARGIKDSTWIGICTISKLDARIQKREQRRRRRASSVLQRRRAQSIERKQES  
EPRIVSRFQCCAWNGGVWFSLLLFYRVFIPVLQSVTARIUGDPSLHGDVWSWLEFFLTSIFSA  
LWVLPFLVLSKVNAIWFQDIADLAFAEYSGRKPFPFSVKIADMLFNLLQALFLIQGMFVSL  
FPHLVGQLVSLHMSLLYSLYCFEYRWFNKGIEHQRLSNERNWPYYFGFCLPLAFLTAMQ  
15 SSYISGCLFSILPFLFIISANEAKTPGKAYLFQLRLFSLVVFLSNRLFKTKTYLQSALSSSTSAEK  
FSPHPSPAKLKATAGH (SEQ ID NO: 637). Accordingly, polypeptides comprising the foregoing amino acid sequence are provided as are polynucleotides encoded such polypeptides.

## 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 115

This gene is expressed primarily in stromal cells and to a lesser extent in multiple sclerosis.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: affecting the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., stromal cells and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of multiple sclerosis and other autoimmune diseases.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 116

This gene is expressed primarily in the gall bladder

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: gall stones or infection of the digestive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system or renal system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for possible prevention of digestive disorders where there may be a lack of digestive enzymes produced or in the detection and possible prevention of gall stones.

## FEATURES OF PROTEIN ENCODED BY GENE NO: 117

The translation product of this gene shares sequence homology with dystrophin gene which is thought to be important in building and maintenance of muscles.

25 This gene is expressed primarily in placenta and to a lesser extent in fetal brain and fetal liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: muscular dystrophy. Duchenne and Becker's muscular dystrophies. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the skeletal muscle system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, brain and other tissue of the nervous system, muscle, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

35

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the dystrophin gene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diseases related the degenerative myopathies that are characterized by the weakness and atrophy of muscles without neural degradation, such as Duchenne and Becker's muscular dystrophies.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 118

This gene is expressed primarily in olfactory tissue and to a lesser extent in cartilage.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: connective tissue diseases: chondrosarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the connective tissue, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue and cartilage, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for tumors of connective tissues, osteoarthritis and the treatment and diagnosis of chondrosarcoma.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 119

This gene is expressed primarily in Activated Neutrophils and to a lesser extent in fetal spleen, and CD34 positive cells from cord blood.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergies, defects in hematopoiesis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential

identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and hematopoiesis system the, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for reducing the allergic effects felt by allergy suffers by neutralizing the activity of the immune system, especially since neutrophils are abundant in persons suffering from allergies and other inflammatory conditions.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 120

The translation product of this gene shares sequence homology with poly A binding protein II which is thought to be important in RNA binding for transcription of RNA to DNA

This gene is expressed primarily in colon and to a lesser extent in brain and immune system.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: colon cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., colon, tissue and cells of the immune system, and brain or other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to poly A binding protein II indicate that polynucleotides and polypeptides corresponding to this gene are useful for detection and treatment of colon cancer and other disorders of the digestive system..

**FEATURES OF PROTEIN ENCODED BY GENE NO: 121**

The translation product of this gene shares sequence homology with thymidine diphosphoglucose 4.6 dehydratase which is thought to be important in the metabolism of sugar.

5 This gene is expressed primarily in fetal liver and spleen and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diabetes. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution and homology to thymidine diphosphoglucose 4.6 dehydratase indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of persons with diabetes since it appears that this protein is needed in the metabolism of sugar in to its more basic components.

**25 FEATURES OF PROTEIN ENCODED BY GENE NO: 122**

The translation product of this gene shares sequence homology with ceruloplasmin which is thought to be important in the metabolism and transport of iron and copper. Ceruloplasmin also contains domains with homology to clotting factors V and VIII. Defects in the circulating levels of ceruloplasmin (aceruloplasminemia) have been associated with certain disease conditions such as Wilson disease, and the accompanying hepatolenticular degeneration.

30 This gene is expressed primarily in brain and retina and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases marked by defects in iron metabolism; aceruloplasminemia not characterized by defects in the

known ceruloplasmin gene locus; nonclassical Wilson disease; movement disorders; and tumors derived from a brain tissue origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, retina, and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, retinal tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution and homology to ceruloplasmin indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of patients with aceruloplasminemia, or other defects in iron and/or copper metabolism. Mutations in this locus could also be diagnostic for patients currently experiencing or predicted to experience aceruloplasminemia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 123**

20 This gene is expressed primarily in brain and B cell lymphoma and to a lesser extent in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: B cell lymphoma; tumors and diseases of the brain and/or spleen; hematopoietic defects. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and hematopoietic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, blood cells, liver, and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

35 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of disorders in neuronal,

hematopoietic, and immune systems. It could potentially be useful for neurodegenerative disorders and neuronal and/or hematopoietic cell survival or proliferation.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 124

This gene is expressed primarily in osteoclastoma, dermatofibrosarcoma, and B cell lymphoma and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer in particular osteoclastoma, dermatofibrosarcoma, and B cell lymphoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, immune, and circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., bone, epidermis, blood cells, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers and lymphoma, osteoporosis; and the control of cell proliferation and/or differentiation.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 125

This gene is expressed primarily in immune tissues and hematopoietic cells, particularly in activated T cells and neutrophils, spleen, and fetal liver, and to a lesser extent in infant adrenal gland.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects in T cell activation; hematopoietic disorders; tumors of a hematopoietic and/or adrenal gland origin. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoietic and/or endocrine systems, expression of this gene at significantly higher

or lower levels may be routinely detected in certain tissues and cell types (e.g., cells and tissues of the immune system, hematopoietic cells, blood cells, liver, and adrenal gland, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune and/or hematopoietic disorders; diseases related to proliferation and/or differentiation of hematopoietic cells; defects in T cell and neutrophil activation and responsiveness; and endocrine and/or metabolic disorders, particularly of early childhood.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 126

This gene is expressed primarily in placenta and endothelial cells and to a lesser extent in melanocytes and embryonic tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell origin; angiogenesis associated with tumor development and metastasis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the vascular system and developing embryo, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, melanocytes, and embryonic tissues, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of developmental disorders; inhibition of angiogenesis; and vascular patterning.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 127

This gene is expressed primarily in endothelial cells and hematopoietic tissues, including spleen, tonsils, leukocytes, and both B- and T-cell lymphomas.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of an endothelial cell and/or hematopoietic origin; leukemias and lymphomas. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and vascular systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endothelial cells, hematopoietic cells, spleen, tonsils, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the manipulation of angiogenesis; the differentiation and morphogenesis of endothelial cells; the proliferation and/or differentiation of hematopoietic cells; and the commitment of hematopoietic cells to distinct cell lineages.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 128

This gene is expressed primarily in kidney medulla and to a lesser extent in spleen from chronic myelogenous leukemia patients, prostate cancer, and some other tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a kidney origin; chronic myelogenous leukemia; prostate cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the kidney and spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., kidney, spleen, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the diagnosis and treatment of kidney disorders and cancer, particularly chronic myelogenous leukemia and prostate cancer. It may also be useful for the enhancement of kidney tubule regeneration in the treatment of acute renal failure.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 129

This gene is expressed primarily in adult and infant brain and to a lesser extent in mesenchymal or fibroblast cells, as well as tissues with a mesenchymal origin.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; fibrosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and of mesenchymal cells and tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for the diagnosis of tumors of a brain and/or mesenchymal origin; neurodegenerative disorders; cancer; and fibrosis, based upon the expression of this gene within those tissues. Fibrosis is considered as mesenchymal cells and fibroblasts are the primary cellular targets involved in this pathological condition.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 130

This gene is expressed primarily in hepatocellular cancer and to a lesser extent in fetal tissues as well as testes tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: liver cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing



immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, fetal tissue, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of liver cancer.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 131

This gene is expressed only in infant early brain.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: development and diseases of the nervous system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

20 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the brain in children and in treating nervous system disorders such as Alzheimer's disease, schizophrenia, dementia, depression, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 132

This gene is expressed primarily in brain and to a lesser extent in glioblastoma.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Alzheimer's disease,

schizophrenia, depression, mania, and dementia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating brain disorders such as Alzheimer's disease, schizophrenia, depression, mania, and dementia.

#### 15 FEATURES OF PROTEIN ENCODED BY GENE NO: 133

The translation product of this gene shares sequence homology with ribitol dehydrogenase of bacteria which is thought to be important in metabolism of sugars.

This gene is expressed primarily in macrophage and to a lesser extent in T-cell lymphoma and lung.

20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tissue destruction in inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and homology to ribitol dehydrogenase indicate that polynucleotides and polypeptides corresponding to this gene are useful for altering macrophage metabolism in diseases such as inflammation where macrophages are causing excess tissue destruction.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 134**

This gene is expressed primarily in pancreatic tumor and to a lesser extent in synovial sarcoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the endocrine and connective tissue systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pancreas, and synovial tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing various cancers.

**20 FEATURES OF PROTEIN ENCODED BY GENE NO: 135**

This gene is expressed primarily in T cell lines such as Raji and to a lesser extent in infant brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune system disorders and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing inflammatory diseases

such as rheumatoid arthritis, sepsis, inflammatory bowel disease, and psoriasis, as well as neutropenia.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 136**

The translation product of this gene shares high sequence homology with SAR1 subfamily of GTP-binding proteins which is thought to be important in vesicular transport in mammalian cells.

This gene is expressed primarily in serum-stimulated smooth muscle cells and to a lesser extent in a T-cell lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: diseases affecting vesicular transport. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the muscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to GTP-binding proteins indicate that polynucleotides and polypeptides corresponding to this gene are useful for gene therapy in treating the large number of diseases involved in defective vesicular transport within cells..

**FEATURES OF PROTEIN ENCODED BY GENE NO: 137**

The translation product of this gene shares sequence homology with a protein found in *C. elegans* cosmid F25B5.

This gene is expressed primarily in a fetal tissues and to a lesser extent in melanocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development, especially of the pulmonary system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes

for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal pulmonary system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissue, pulmonary tissue, and melanocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

10 corresponding to this gene are useful for treatment and diagnosis of diseases affecting the pulmonary system, such as emphysema.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 138

This gene is expressed primarily in gall bladder and to a lesser extent in smooth muscle.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: digestive system disease and gall bladder problems. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the digestive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., gall bladder and tissue of the digestive system, and smooth muscle, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

30 corresponding to this gene are useful for treating diseases of the digestive system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 139

This gene is expressed primarily in placenta and to a lesser extent in brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal fetal development. Similarly, polypeptides and antibodies directed to these polypeptides are

useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of developing tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating and diagnosing abnormal fetal development.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 140

This gene is expressed primarily in smooth muscle and to a lesser extent in ovary, prostate cancer, and activated monocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hypertension and atherosclerosis. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., smooth muscle, ovary and other reproductive tissue, prostate, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating diseases of the circulatory system, such as hypertension, atherosclerosis, etc.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 141

This gene is expressed primarily in fetal spleen and to a lesser extent in placenta and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and other diseases affecting blood cells. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the circulatory and pulmonary systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen, placenta, bone marrow, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the generation of red and white blood cells and for the diagnosis of disease of these cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 142

The predicted translation product of this contig is a human homolog of the murine tetracycline/sugar transporter molecule recently reported by Matsu and colleagues (Biochem. Biophys. Res. Commun. 238 (1), 126-129 (1997)).

This gene is expressed primarily in synovium and to a lesser extent in endothelial cells.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: rheumatoid arthritis and inflammation. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and lymphatic systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, and endothelial cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment and diagnosis of inflammatory diseases, such as rheumatoid arthritis, leukemia, neutropenia, inflammatory bowel disease, psoriasis, sepsis, and the like.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 143

This gene is expressed primarily in placenta and to a lesser extent in melanocyte, fetal liver and spleen, and bone marrow.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: abnormal early development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, melanocytes, liver, spleen, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and diagnosis of abnormal early development phenomena and diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 144

This gene is expressed primarily in fetal liver and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: anemia and neutropenia. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and blood systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver and spleen, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the

expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful in hematopoiesis and bone marrow regeneration as it is most abundant in fetal tissues responsible for the generation of hematopoietic cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 145

The translation product of this gene shares sequence homology with protein tyrosine phosphatase which is thought to be important in transducing signal to activate cells such as T cell, B cell and other cell types.

This gene is expressed primarily in T cells and tissues in early stages of development and to a lesser extent in cancers.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic and fetal tissue, undifferentiated cells, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the protein tyrosine phosphatase family indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating the immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 146

This gene is expressed primarily in T cell and to a lesser extent in B cell, macrophages and tumor tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in

providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating the immune system therefore can be used in treating diseases such as autoimmune diseases and cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 147

This gene is expressed primarily in placenta and to a lesser extent in endothelial cells, testis tumor, ovarian cancer, uterine cancer.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, endothelial cells, testis and ovary and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 148

This sequence has significant homology to mouse torsin A. Recently, another group cloned the human Torsin A gene. (See, Accession No. 2358279; see also Nature Genet. 17, 40-48 (1997).)

This gene is expressed primarily in osteoclastoma, T-cell, and placenta and to a lesser extent in fetal liver, fetal brain, adult brain and tumor tissues

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: disease conditions in hematopoiesis and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, bone, placenta, lung, liver, and brain and other tissues of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treating blood related diseases such as deficiencies in red blood cell, white blood cell, platelet and other hematopoiesis cells.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 149

This gene is expressed primarily in T cell, prostate and prostate cancer, endothelial cells and to a lesser extent in monocyte, dendritic cell, bone marrow, salivary gland, colon cancer, stomach cancer, pancreatic tumor, uterine cancer, fetal spleen and osteoclastoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immuno-related diseases and cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, prostate, endothelial cells, dendritic cells, bone marrow, salivary gland, colon, stomach, pancreas, uterus, spleen and bone, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 150

This gene was recently cloned by another group, calling it cIF3-p66. (See Accession No. 2351378.) This gene plays a role in RNA binding and macromolecular assembly, and therefore, any mutations in this gene would likely result in a diseased phenotype. Preferred polypeptide fragments comprise the amino acid sequence:

MAKFMTPVIQDNPNSGWCPCAVPEQFRDMPYQPFSGKDRLGKLVADWTOATTQDKRYTNKYSS  
QFGGGSQAYFHEEDESSEFQLVDTARTQKTAQQRNMRFAQRNLRDRKDRRNMLQPNLQILP  
KSAKQKERERLRLQKKFKQFQVRQKWDQSKQKPRDSSVEVRSDWVYKEEMDPQLMKM  
LEVSEPDIECCGGALEYDYDKAFDRITRSEKPLRXKKRIFHTVTTDDPVIRKLAKTQGNVFA  
AILATLMCTRSYVSWDITVQVRVGSKLFFDKRDNDFDLTLVSETANEPQDEGNSRNPRL  
AMEATYNHNSQQCLRMGKERYNFPNPFVEDDDMDKNEIASVAYRSGKLGGDIDLTVRC  
EHDGVMTOANGEVSEFINIKTLNWDNRHNGVDWRQKLDQRGAVIATELKNNSYKLARWTC  
CALLAGSEYLKLOYSVRYHVKDSRRHVLGTQQFKPNEFASQINLSVENAWGLRCVIDICMIL  
EEOGYLLKDPNKQVIRVYSLPDGTFSS (SEQ ID NO: 638), as well as N-terminal and C-terminal deletions of this polypeptide fragment.

This gene is expressed primarily in T cell, bone marrow, embryo and endothelial cells and to a lesser extent in testis tumor and endometrial tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune diseases and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for immune disorders and cancers.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 151**

This gene is expressed primarily in testis and to a lesser extent in T cell, spinal cord, placenta, neutrophil and monocyte.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: male reproductive and endocrine disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive, immune and endocrine systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testis and other reproductive tissue, blood cells, tissue of the nervous system, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for regulating immune and reproductive functions.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 152**

The translation product of this gene shares sequence homology with tyrosyl-tRNA synthetase which is thought to be important in cell growth.

This gene is expressed primarily in brain, liver, keratinocytes, tonsils, and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer autoimmune diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, keratinocytes, tonsils, heart expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissues of the nervous system, liver, keratinocytes, tonsils and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard

gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to tyrosyl-tRNA synthetase indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 153**

This gene is homologous to the *Drosophila* transcriptional regulator *dre4*. (See Accession No. 2511745.) *Dre4* is a gene required for steroidogenesis in *Drosophila melanogaster* and encodes a developmentally expressed homologue of the yeast transcriptional regulator CDC68. Preferred polypeptide fragments comprise the amino acid sequence: KKRRTDVQFYTEVGEITTDLGKQHQMHDRDDLVAEQMERMRHKLKTAFLKFIKVEALTKEELEPEVPRDLGFNGAPYRSTCLLQPTSSALVNA TEWPPFVVTLDVELIHFXR VQFHLKNFDMVIVYKDYSSKYVTMINAIPVASLDPIKEWLNSCDLKYTEGVQSLNWTGMKTIVD DPEQFTEQCGWSFL (SEQ ID NO: 639), as well as N-terminal and C-terminal deletions of this fragments. Also preferred are polynucleotide fragments encoding this polypeptide fragment.

This gene is expressed primarily in fetal liver, spleen, placenta, lung, T cell, thyroid, testes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain tumor, heart and liver diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the fetal liver, spleen, placenta, lung, T cell, thyroid, testes expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, spleen, placenta, lung, blood cells, thyroid, and testes and other reproductive tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 154**

This gene is expressed primarily in brain and to a lesser extent in fetal heart, testis, spleen, lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: heart, liver and spleen diseases, immunological diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, fetal heart, testis, spleen, lung expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, heart, testes and other reproductive tissue, spleen, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 155

Activation of T cells through the T cell antigen receptor (TCR) results in the rapid tyrosine phosphorylation of a number of cellular proteins, one of the earliest being a 100 kDa protein. This gene is the human equivalent of murine valosin containing protein (VCP). VCP is a member of a family of ATP binding, homo-oligomeric proteins, and the mammalian homolog of *Saccharomyces cerevisiae* cdc48p, a protein essential to the completion of mitosis in yeast. Both endogenous and expressed murine VCP are tyrosine phosphorylated in response to T cell activation. Thus we have identified a novel component of the TCR mediated tyrosine kinase activation pathway that may provide a link between TCR activation and cell cycle control.

This gene is expressed primarily in brain, liver, spleen, placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, spleen, placenta expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, spleen, and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from

35

an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to VCR indicate that polynucleotides and polypeptides corresponding to this gene are useful for treating cancer.

5

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 156

The translation product of this gene shares sequence homology with rat growth response protein which is thought to be important in cell growth. A group recently cloned the human homolog of this gene, calling it insulin induced protein 1. (See Accession No. 2358269, see also, Genomics 43 (3), 278-284 (1997).) Preferred polypeptide fragments comprise the amino acid sequence: RSLGLGDTIAFLATLTQF LVYNGVYQYSPDFLYIRSWLPCIFFSGVTVGNIGROLAMGVPEKPHSD (SEQ ID NO: 640), as well as N-terminal and C-terminal deletions of this polypeptide fragment. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

15

This gene is expressed primarily in brain, liver, placenta, heart, spleen, lymphoma.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain, liver, placenta, heart, spleen, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, placenta, heart, spleen, and lymphoid tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25

The tissue distribution and homology to growth-response protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for modulating cell growth.

30

35



**FEATURES OF PROTEIN ENCODED BY GENE NO: 157**

This gene is expressed primarily in Glioblastoma, endometrial tumor, lymphoma and pancreas tumor.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Glioblastoma, Endometrial tumor, lymphoma and pancreas tumor. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, lymphoid tissue, pancreas, and tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 158**

The translation product of this gene shares sequence homology with IGE receptor which is thought to be important in allergy and asthma.

This gene is expressed primarily in T cell, and fetal liver.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: allergy and asthma and other immunological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to IgE receptor indicate that

polynucleotides and polypeptides corresponding to this gene are useful for allergy and asthma.

**5 FEATURES OF PROTEIN ENCODED BY GENE NO: 159**

The translation product of this gene shares sequence homology with immunoglobulin heavy chain which is thought to be important in immune response to the antigen.

This gene is expressed primarily in activated neutrophil and to a lesser extent in activated T cell, monocyte and heart.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: infection, inflammation and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, and heart, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to immunoglobulin heavy chain variable region indicate that polynucleotides and polypeptides corresponding to this gene are useful for making the ligand to block specific antigen which cause certain disease.

**FEATURES OF PROTEIN ENCODED BY GENE NO: 160**

The translation product of this gene shares sequence homology with mouse X inactive specific transcript protein which is thought to be important in X chromosome inactivation.

This gene is expressed primarily in HSA172 cell and to a lesser extent in normal ovary tissue, ovarian cancer, frontal cortex and brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ovarian tumor, schizophrenia and other neurological disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for

differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., ovary and other reproductive tissue, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to X inactive specific transcript protein indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of reproductive system tumors and CNS tumors.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 161

This gene is expressed primarily in adipose cell and to a lesser extent in liver and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: obesity and liver disorder. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the adipose cell, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., adipose cells, liver, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for treatment of obesity and liver disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 162

The translation product of this gene shares sequence homology with yeast ubiquitin activating enzyme homolog which is thought to be important in protein posttranslation processing.

This gene is expressed primarily in stromal cell and to a lesser extent in retina, H. Atrophic Endometrium, colon carcinoma and myeloid progenitor cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: defects of stromal cell development, neuronal growth disorders and tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., retinal cells, endometrium, colon, and bone marrow, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ubiquitin-activating enzyme homolog indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis or treatment of some type of tumors, fucosidosis and neuronal growth disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 163

This gene is expressed primarily in primary breast cancer and hemangiopericytoma and to a lesser extent in adult brain and cerebellum.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer, leukemia and cerebellum disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system and neural system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides

corresponding to this gene are useful for diagnosis or treatment of various tumors and disease involved in neural system.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 164

The translation product of this gene shares sequence homology with proline rich proteins. Recently, another group has also cloned this gene, calling it CD84 leukocyte antigen, a new member of the Ig superfamily. (See Accession No. U82988, see also, Blood 90 (6), 2398-2405 (1997).)

10 This gene is expressed primarily in Weizmann olfactory tissue and osteoclastoma and to a lesser extent in anergic T-cell.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: ostitis and immune disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., olfactory tissue, bone, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

25 The tissue distribution and homology to the Ig superfamily indicate that the protein product of this clone is useful for treatment of osteoporosis, autoimmune disease, and other immune disorders.

#### 30 FEATURES OF PROTEIN ENCODED BY GENE NO: 165

This gene is expressed primarily in atrophic endometrium and colon cancer and to a lesser extent in some fetal tissues.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: tumors. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system,

expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., endometrium, colon, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of tumors, specifically endometrium and colon tumors.

#### 10 FEATURES OF PROTEIN ENCODED BY GENE NO: 166

This gene is expressed primarily in human primary breast cancer and to a lesser extent in activated monocyte. Although the predicted signal sequence is identified in Table 1, other upstream sequences are also relevant. Preferred polypeptide fragments comprise the amino acid sequence: VTQPKHLASMGGSVEIFSFYYPWELAXXPXVRLSWRRGHFHG QSFYSTRPPSHKDYNNRLFLNWTGQESGFLRISNLRKEDQSVYPCRVELDTRSG (SEQ ID NO: 641), as well as N-terminal and C-terminal deletions. Also preferred are polynucleotide fragments encoding these polypeptide fragments.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: breast cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., mammary tissue, and blood cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis of breast cancer.

#### 35 FEATURES OF PROTEIN ENCODED BY GENE NO: 167

This gene is expressed primarily in fetal tissues and to a lesser extent in adult lung. This gene has also been mapped to chromosomal location 9q34, and thus, can be used as a marker for linkage analysis for chromosome 9.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the embryo tissues, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., fetal tissues, and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 168

The translation product of this gene shares sequence homology with Ig Heavy Chain which is thought to be important in immune response.

This gene is expressed primarily in prostate cancer tissue specifically

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: prostate cancer.

Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate, tissue and cells of the immune system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 169

The translation product of this gene shares sequence homology with cytosolic acyl coenzyme-A hydrolase, which is thought to be important in neuron-specific fatty acid metabolism. The gene represented by this contig has since been published by Hajra and colleagues (GenBank Accession No. U91316).

This gene is expressed primarily in human pituitary gland and to a lesser extent in colorectal cancer tissue. This gene has also been observed in the LNCAP cell line.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: hyperlipidemias of familial and/or idiopathic origins. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly blood, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., pituitary and colon, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to rat cytosolic acyl coenzyme-A

hydrolase indicate that polynucleotides and polypeptides corresponding to this gene are useful for the detection or treatment of hyperlipidemia disease states by virtue of the ability of specific drugs to activate the enzyme.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 170

The translation product of this gene shares sequence homology with a *Caenorhabditis elegans* gene which is thought to be important in organism development.

This gene is expressed primarily in human synovial sarcoma tissue, bone marrow, and to a lesser extent in human brain.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, of bone, specifically synovial sarcoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the bone, connective tissues and possibly immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., synovial tissue, bone marrow, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another

tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

- 5 The tissue distribution and homology to *Caenorhabditis elegans* indicate that polynucleotides and polypeptides corresponding to this gene are useful as a diagnostic and/or therapeutic modality directed at the detection and/or treatment of connective tissue sarcomas or other related bone diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 171

- 10 The translation product of this gene shares sequence homology with beta1-6GlcNAc transferase which is thought to be important in the transfer and metabolism of beta1-6, N-acetylglucosamine. This gene product has previously been shown to suppress melanoma lung metastasis in both syngeneic and nude mice, decreased invasiveness into the matrigel, and inhibition of cell attachment to collagen and laminin without affecting cell growth.

15 This gene is expressed primarily in human testes and prostate tissues, and to a lesser extent in kidney, medulla, and pancreas.

- 20 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancer particularly melanoma. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., testes and other reproductive tissue, prostate, kidney, pancreas, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30 The tissue distribution and homology to beta1-6GlcNAc transferase indicate that the protein product of this clone is useful for the development of diagnostic and/or therapeutic modalities directed at the detection and/or treatment of cancer, the metastasis of malignant tissue or cells. Defects in this potentially secreted enzyme may play a role in metastasis.

35

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 172

This gene is expressed primarily in fetal spleen and liver.

- 5 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: immune disorders, Wilm's tumor disease, hepatic disorders, and hematopoietic disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the hematopoiesis and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., spleen and liver, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

15 The tissue distribution indicates that polynucleotides and polypeptides corresponding to this gene are useful for the treatment and identification of fetal defects along with correcting diseases that affect hematopoiesis and the immune system.

20

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 173

The translation product of this gene shares sequence homology with ret II oncogene which is thought to be important in Hirschsprung disease and many types of cancers.

- 25 This gene is expressed in multiple tissues including the lymphatic system, brain, and thyroid.

30 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: Hirschsprung disease and multiple cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune and central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., lymphoid tissue, thyroid, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to

the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to ret II oncogene indicate that polynucleotides and polypeptides corresponding to this gene are useful for diagnosis and treatment of various cancers. It would also be useful for the diagnosis and treatment of Hirschsprung disease. Preferred polypeptides of the invention comprise the amino acid sequence: MEAQQVNEAESAREQLQXLHDQIAGQKASKQELETERLKRQEFHYTEEDLY RTKNTLSRIKDRDEIQKLRNQLTNKTLSNSQSELENRLHQLJETILQKQTMLESISTEKNLS VFQLERLEQQMNSAGSSSSNGSINMSGIDNCGEGLRNVPVLFNDTETNLNAGMYGKYRKAAS SIDQFSIRLGIPLRRYPPIARVFVHYIMALLHLWYVMIVLLTYPTEM HHDPQYCK (SEQ ID NO: 642).

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 174

The translation product of this gene shares sequence homology with testis enhanced gene transcript which is thought to be important in regulation of human development.

This gene is expressed primarily in infant brain and to a lesser extent in a variety of other tissues and cell types, including the prostate, testes, monocytes, macrophages, dendritic cells, keratinocytes, and adipocytes.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neurological, developmental, immune and inflammation disorders. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the brain and immune systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, prostate, testes and other reproductive tissue, blood cells, keratinocytes, and adipocytes, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to testis enhanced gene transcript indicate that the protein product of this clone is useful for diagnosis and treatment of disorders involving the developing brain and the immune system.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 175

This gene is expressed primarily in prostate and to a lesser extent in various other tissues, including placenta.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, cancers, especially of the prostate. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the prostate, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., prostate and placenta, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone is useful for diagnosis and treatment of prostate disorders and cancer. It may also be useful for the diagnosis and treatment of endocrine disorders.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 176

The translation product of this gene shares sequence homology with *Saccharomyces cerevisiae* YNT20 gene which is thought to be important in mitochondrial function.

This gene is expressed at a particularly high level in muscle tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases related to such tissues and cell types including: muscle wasting diseases. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., muscle and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e.,

the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to the YNT20 gene indicate that this protein is useful for treatment and detection of neuromuscular diseases caused by loss of mitochondrial function. For example this gene or its protein product could be used in replacement therapy for such diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 177

This gene is expressed primarily in the brain and to a lesser extent in kidney, placenta, smooth muscle, heart and lung.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: neuromuscular diseases, degenerative diseases of the central nervous system, and heart disease. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the neuromuscular system, central nervous system, and heart, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, kidney, placenta, muscle, heart and lung, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

This gene or its protein product could also be used for replacement therapy for the above mentioned diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 178

The translation product of this gene shares sequence homology with caldesmon which is thought to be important in the cellular response to changes in glucose levels.

This gene is expressed primarily in multiple tissues including brain and retina. Therefore, polynucleotides and polypeptides of the invention are useful as reagents for identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: central nervous system disorders and retinopathy. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the CNS disorders and retinopathy, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, and retinal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to caldesmon indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of retinopathies.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 179

The translation product of this gene shares sequence homology with mouse fibrosin protein which is thought to be important in regulation of fibrinogenesis in certain chronic inflammatory diseases.

This gene is expressed primarily in amniotic cells and breast tissue.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of breast cancer and abnormal embryo development. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., amniotic cells, and mammary tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution and homology to fibrosin indicate that the protein product of this clone is useful for treatment of breast cancer. This gene or its protein product could be used in replacement therapy for breast cancer. In addition the protein product of this gene is useful in the treatment of chronic inflammatory diseases.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 180

This gene is expressed several infant tissues including brain and liver and various adult tissues including brain, lung, liver, testes, and prostate.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions which include, but are not limited to, brain cancer, lung cancer, liver cancer and cancers of the reproductive system. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, hepatic system, and reproductive system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, liver, testes and other reproductive tissue, and prostate, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution of this gene product indicates that the protein product of this clone is involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### 20 FEATURES OF PROTEIN ENCODED BY GENE NO: 181

This gene is expressed primarily in activated monocytes and to a lesser extent in melanocytes and dendritic cells.

Therefore, polynucleotides and polypeptides of the invention are useful as

reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of immune system diseases and cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., blood cells, melanocytes, and dendritic cells, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The tissue distribution indicates that the protein product of this clone could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### 5 FEATURES OF PROTEIN ENCODED BY GENE NO: 182

This gene is expressed primarily in placenta and several tumors of various tissue origin and to a lesser extent in normal tissues including liver, lung, brain, and skin,

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of cancers of all kinds. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, respiratory system and skin, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., liver, lung, brain and other tissues of the nervous system, and skin, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The high expression of this gene in multiple tumors indicates that the protein product of the clone may be involved in cell growth control and therefore would be useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

#### 25 FEATURES OF PROTEIN ENCODED BY GENE NO: 183

The translation product of this gene shares sequence homology with the mouse Ndr1 gene which is thought to be important in cancer progression.

This gene is expressed multiple cell types and tissues including brain, lung, kidney, bone marrow, liver, and spleen.

Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the nervous, immune, and endocrine

systems, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, lung, kidney, bone marrow, liver and spleen, and cancerous and wounded



(tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

5 The tissue distribution and homology to Ndr1 gene, which is thought to be involved in cancer progression, indicate that polynucleotides and polypeptides corresponding to this gene are useful for treatment of certain cancers. Likewise molecules developed to block the activity of the protein product of this clone could be used to block its potential role in tumor growth promotion.

10

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 184

This gene is expressed primarily in early stage human brain and liver and to a lesser extent in several other fetal tissues.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of diseases and conditions: brain and liver cancers. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system and immune system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., brain and other tissue of the nervous system, liver, and fetal tissue, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

30

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 185

This gene is expressed primarily in infant and embryonic brain.

35 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of degenerative nervous system disorders and brain cancer. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell

type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., embryonic tissue, brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

10 The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

#### FEATURES OF PROTEIN ENCODED BY GENE NO: 186

This gene is expressed primarily in multiple tissues including placenta, fetal lung, fetal liver, and brain.

15 Therefore, polynucleotides and polypeptides of the invention are useful as reagents for differential identification of the tissue(s) or cell type(s) present in a biological sample and for diagnosis of all types of cancers including liver, brain and lung. Similarly, polypeptides and antibodies directed to these polypeptides are useful in providing immunological probes for differential identification of the tissue(s) or cell type(s). For a number of disorders of the above tissues or cells, particularly of the central nervous system, pulmonary system, and hepatic system, expression of this gene at significantly higher or lower levels may be routinely detected in certain tissues and cell types (e.g., placenta, lung, liver, and brain and other tissue of the nervous system, and cancerous and wounded tissues) or bodily fluids (e.g., serum, plasma, urine, synovial fluid or spinal fluid) or another tissue or cell sample taken from an individual having such a disorder, relative to the standard gene expression level, i.e., the expression level in healthy tissue or bodily fluid from an individual not having the disorder.

30

The expression of this gene in embryonic tissues indicates that the protein could be involved in growth regulation and could be used as a growth factor or growth blocker in a variety of settings including treatment of cancers.

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	11	582	1	582	177	177	313	1	18	19	22
1	HTTEZ21	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	197	1020	296	830	442	442	499	1	18	19	22
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	12	465	1	465	81	81	314	1	30	31	128
2	HBGBW52	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	198	524	229	343		196	500	1	20	21	33
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	13	474	1	474	1	1	315	1	24	25	28
3	HCUFM41	97897 02/26/97 209043 05/15/97	ZAP Express	199	332	1	319	35	35	501	1	24	25	28

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
4	HCUFQ22	97897 02/26/97 209043 05/15/97	ZAP Express	14	314	1	298	122	122	316	1	34	35	64
5	HCUFV01	97897 02/26/97 209043 05/15/97	ZAP Express	15	613	1	613	30	30	317	1	18	19	21
6	HCUGA50	97897 02/26/97 209043 05/15/97	ZAP Express	16	356	1	356	239	239	318	1	22	23	39
7	HCUIM14	97897 02/26/97 209043 05/15/97	ZAP Express	17	414	185	414	278	278	319	1	26	27	33
8	HLDOU93	97897 02/26/97 209043 05/15/97	pCMVSPORT 3.0	18	469	1	469	77	77	320	1	44	45	88
9	HEIAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	19	550	1	550	129	129	321	1	21	22	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
9	HEIAX07	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	200	376	9	376		1	502	1	8	9	15
10	HSAXR76	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	20	741	55	741	190	190	322	1			27
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	21	991	1	991	62	62	323	1	30	31	64
11	HNGJJ68	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	201	1192	253	1137		409	503	1			19
12	HCFAW04	97897 02/26/97 209043 05/15/97	pSport1	22	653	1	653	64	64	324	1	30	31	196
12	HCFAW04	97897 02/26/97 209043 05/15/97	pSport1	202	589	1	513	109	109	504	1			29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	23	1486	596	1418	102	102	325	1	54	55	252
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	203	847	1	839	87	87	505	1	30	31	75
13	HLMAV65	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	204	852	75	850		690	506	1			10
13	HTXEF04	209235 09/04/97	Uni-ZAP XR	205	1354	54	1354	100	100	507	1	33	34	207
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	24	2323	1017	2059	1242	1242	326	1	21	22	68
14	HPMFD84	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	206	1378	113	1226	303	303	508	1	25	26	36
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	25	683	1	683	304	304	327	1	30	31	84

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
15	HE6DB26	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	207	1166	281	884	567	567	509	1	18	19	19
16	HHFFL33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	26	2036	14	1959	214	214	328	1	20	21	36
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	27	717	1	717	70	70	329	1	30	31	63
17	HODBD33	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	208	697	2	697	33	33	510	1	31	32	32
18	HMDAE90	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	28	495	1	495	39	39	330	1	24	25	35
19	HOUAW01	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	29	556	1	556	116	116	331	1	19	20	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
20	HBJAE44	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	30	434	1	434	78	78	332	1	35	36	40
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	31	715	1	715	87	87	333	1	30	31	111
21	HCFME41	97897 02/26/97 209043 05/15/97	pSport1	209	932	274	932	387	387	511	1	27	28	28
22	HOGCO71	97897 02/26/97 209043 05/15/97	pCMVSport 2.0	32	486	1	486	137	137	334	1	21	22	106
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	33	725	1	725	436	436	335	1	30	31	50
23	HOSEX08	97897 02/26/97 209043 05/15/97	Uni-ZAP XR	210	661	1	647	81	81	512	1	25	26	26

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
24	HSKNJ72	97897 02/26/97 209043 05/15/97	pBluescript	34	437	1	437	85	85	336	1	30	31	48
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	35	943	1	943	196	196	337	1	30	31	41
25	HEBEB69	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	211	592	1	534	72	72	513	1	24	25	33
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	36	604	1	604	375	375	338	1	20	21	76
26	HE6EH18	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	212	938	1	509		17	514	1	30	31	47
27	HSAUZ47	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	37	349	1	349		201	339	1	20	21	31

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
28	HSSDM73	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	38	672	1	672	22	22	340	1	38	39	42
29	HBMVK68	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	39	1908	135	1908	309	309	341	1	20	21	26
30	HMKDC66	97898 02/26/97 209044 05/15/97	pSport1	40	458	93	458	147	147	342	1	24	25	26
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	41	1153	500	1153	427	427	343	1	30	31	157
31	HMKCU94	97898 02/26/97 209044 05/15/97	pSport1	213	1079	502	896		739	515	1	23	24	43
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	42	1983	1092	1983	27	27	344	1	11	12	520

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
32	HRDEW41	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	214	3791	2757	3357		2030	516	1			3
33	HTOJN06	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	43	1406	1	695		19	345	1	19	20	39
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	44	1391	851	1153	74	74	346	1	30	31	234
34	HBGDA21	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	215	1334	822	1036		638	517	1	18	19	174
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	45	1569	768	1569	14	14	347	1	19	20	169
35	HFGAK75	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	216	1511	770	1404	844	844	518	1	32	33	43

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
36	HHPBD40	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	46	1924	1	1681	62	62	348	1	19	20	43
37	HOVCL83	97898 02/26/97 209044 05/15/97	pSport1	47	475	252	396	141	141	349	1	37	38	78
38	HBCAY62	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	48	346	1	346	61	61	350	1	19	20	24
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	49	1366	882	1300	177	177	351	1	30	31	274
39	HBICM48	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	217	642	192	581		448	519	1			13
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	50	1405	110	1404	61	61	352	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
40	HLTCL35	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	218	1241	1	1241	172	172	520	1	21	22	30
41	HLHCK50	97898 02/26/97 209044 05/15/97	Uni-ZAP XR	51	504	207	485	222	222	353	1			3
42	HRSAN45	97899 02/26/97 209045 05/15/97	ZAP Express	52	777	1	214	113	113	354	1	24	25	52
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	53	602	1	419	41	41	355	1	59	60	132
43	HSNBB14	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	219	1080	186	686	399	399	521	1	26	27	47
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	54	1749	222	1749	166	166	356	1	30	31	204

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
44	HMABL38	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	220	1258	149	1190	254	254	522	1	18	19	26
45	HSKDK47	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	55	1896	596	1614	650	650	357	1	33	34	47
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	56	1753	555	1753	414	414	358	1	18	19	73
46	HOSFH03	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	221	1693	554	1693		526	523	1	25	26	58
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	57	1220	690	1024	128	128	359	1	30	31	102
47	HOGAV75	97899 02/26/97 209045 05/15/97	pCMVSPORT 2.0	222	1196	712	1163		1097	524	1			19

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
48	HFCA174	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	58	1049	362	1049	335	335	360	1	33	34	48
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	59	1776	854	1737	189	189	361	1	30	31	179
49	HAGBI17	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	223	1791	979	1791	1164	1164	525	1	18	19	40
50	HLFBC91	97899 02/26/97 209045 05/15/97	pBluescript SK-	60	443	1	443	164	164	362	1	21	22	25
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	61	2888	1909	2888	90	90	363	1	30	31	224
51	HPRCA31	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	224	2517	1597	2517	1953	1953	526	1	18	19	57

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	62	1851	1568	1736	139	139	364	1	30	31	349
52	HPRCE95	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	225	2424	299	2309		530	527	1	17	18	21
53	HHTLC66	97899 02/26/97 209045 05/15/97	ZAP Express	63	3542	883	3492	964	964	365	1	25	26	467
54	HMADJ02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	64	883	237	883	229	229	366	1	30	31	152
54	HMADJ02	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	226	1080	242	1033	436	436	528	1	24	25	39
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	65	1541	1	1541	236	236	367	1	30	31	373



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
55	HPRCU93	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	227	1336	4	1336	946	946	529	1	25	26	128
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	66	732	41	698	163	163	368	1	18	19	83
56	HSAXS65	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	228	2043	1133	1756	1262	1262	530	1	20	21	82
57	HKTAG35	209011 04/28/97	Uni-ZAP XR	67	629	1	629	264	264	369	1			21
57	HMEFX42	97899 02/26/97 209045 05/15/97	Lambda ZAP II	229	540	25	536	227	227	531	1			20
58	HHFHN61	97899 02/26/97 209045 05/15/97	Uni-ZAP XR	68	1751	375	1751	95	95	370	1	19	20	227
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	69	508	1	508	22	22	371	1	30	31	79

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
59	HCWEF90	97899 02/26/97 209045 05/15/97	ZAP Express	230	448	9	448		1	532	1	22	23	75
60	HHGCM20	97899 02/26/97 209045 05/15/97	Lambda ZAP II	70	245	1	245	93	93	372	1	1	2	51
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	71	361	1	361	1	1	373	1	30	31	61
61	HFRAU10	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	231	407	1	407	210	210	533	1	17	18	60
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	72	713	8	713	169	169	374	1	30	31	40
62	HATDT67	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	232	830	190	580	329	329	534	1	28	29	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	73	862	1	862	67	67	375	1	30	31	44
63	HOUBG93	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	233	932	138	905	287	287	535	1			2
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	74	4602	4162	4525	730	730	376	1	30	31	203
64	HMWEX24	97900 02/26/97 209046 05/15/97	Uni-Zap XR	234	2786	2406	2739	2577	2577	536	1	22	23	36
65	HSGBA84	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	75	1255	1	1195	112	112	377	1	28	29	29
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	76	475	1	475	13	13	378	1	30	31	136

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
66	HTOCD52	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	235	458	1	458	26	26	537	1			14
67	HTGCP16	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	77	465	25	299	74	74	379	1	33	34	41
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	78	1907	1627	1730	26	26	380	1	30	31	468
68	HKIXR69	97900 02/26/97 209046 05/15/97	pBluescript	236	591	1	444	251	251	538	1			18
69	HETGJ09	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	79	1168	136	1168	267	267	381	1	20	21	29
70	HOBNC61	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	80	1285	132	1285	292	292	382	1	27	28	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
71	HFFAH94	97900 02/26/97 209046 05/15/97	Lambda ZAP II	81	1290	768	1054	701	701	383	1	21	22	138
72	HBLAI95	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	82	684	1	684	119	119	384	1	30	31	74
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	83	2024	1609	1953	200	200	385	1	30	31	521
73	HSQEL25	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	237	1286	391	959		1204	539	1	9	10	11
74	HEBEG68	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	84	931	14	537	85	85	386	1	25	26	137
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	85	825	59	802	66	66	387	1	30	31	186

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	238	734	1	734	1	1	540	1	37	38	108
75	HBLAB39	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	239	809	80	794		294	541	1	15	16	106
76	HTXDU73	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	86	1238	36	918	17	17	388	1			1
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	87	1460	9	1458	166	166	389	1	53	54	299
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	240	2201	841	2080	507	507	542	1	43	44	136
77	HOEAS24	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	241	1661	311	1520	390	390	543	1	35	36	424

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
78	HTEIY30	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	88	1395	567	1395	639	639	390	1	36	37	49
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	89	1186	352	1186	540	540	391	1	49	50	61
79	HSKNE46	97900 02/26/97 209046 05/15/97	pBluescript	242	1146	329	1146	564	564	544	1	21	22	39
80	HPMFL27	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	90	1821	1203	1614	1503	1503	392	1	30	31	79
81	HMWDN32	97900 02/26/97 209046 05/15/97	Uni-Zap XR	91	862	253	862	359	359	393	1	32	33	36
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	92	696	349	696	98	98	394	1	30	31	180

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
82	HPRAX55	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	243	1350	265	1230	348	348	545	1	32	33	58
83	HHFFW36	97900 02/26/97 209046 05/15/97	Uni-ZAP XR	93	1886	1	1759	197	197	395	1			21
84	HE2PL77	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	94	1774	742	1772	785	785	396	1	21	22	60
85	HSDFV29	209076 05/22/97	Uni-ZAP XR	95	2503	1	1648	206	206	397	1	32	33	152
85	HCQAV53	97901 02/26/97 209047 05/15/97	Lambda ZAP II	244	1529	72	911	191	191	546	1	20	21	33
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	96	2801	418	2801	234	234	398	1	30	31	480
86	HTPEG42	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	245	1537	1	1537	125	125	547	1	21	22	367

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
87	HLHDR57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	97	1631	916	1631	1	1	399	1	1	2	423
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	98	504	26	504	197	197	400	1	23	24	78
88	HAUAV32	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	246	506	1	499	183	183	548	1	32	33	77
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	99	1416	145	1416	456	456	401	1	18	19	74
89	HNEBI60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	247	1348	84	1348	363	363	549	1	21	22	47
90	HSJC116	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	100	2847	1	2847		2	402	1			20

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
91	HTSEL31	97901 02/26/97 209047 05/15/97	pBluescript	101	1394	608	1346	602	602	403	1	23	24	87
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	102	794	1	794	518	518	404	1	30	31	92
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	248	1766	42	1766	356	356	550	1	30	31	168
92	HAUBL57	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	249	2664	47	1708		147	551	1	18	19	124
93	HODAS59	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	103	1544	898	1531	975	975	405	1			21
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	104	871	106	871	248	248	406	1	34	35	174

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
94	HE6CT48	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	250	865	97	865	258	258	552	1	19	20	177
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	105	404	1	404	16	16	407	1	21	22	64
95	HMDAA61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	251	2082	852	2074	829	829	553	1	22	23	72
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	106	1542	506	1542	122	122	408	1	51	52	280
96	HAQBK61	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	252	1482	508	1482		633	554	1	15	16	45
96	HCUHB01	209215 08/21/97	ZAP Express	253	834	1	834	82	82	555	1	40	41	251
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	107	2327	1528	2327	465	465	409	1	30	31	284

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
97	HAQBF73	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	254	1508	885	1508		988	556	1			19
98	HAQBT94	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	108	1062	157	1062	172	172	410	1	28	29	187
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	109	2539	275	2501	903	903	411	1	30	31	237
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	255	2514	592	2431	176	176	557	1	30	31	217
99	HETHE07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	256	2357	465	2288		1151	558	1	12	13	82
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	110	1751	969	1751	4	4	412	1	46	47	192

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
100	HLQAB52	97901 02/26/97 209047 05/15/97	Lambda ZAP II	257	689	218	655	314	314	559	1	18	19	95
100	HEONN58	209119 06/12/97	pSportI	258	2377	5	2377	25	25	560	1	28	29	54
101	HCRAM28	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	111	1117	1	1117		1	413	1	19	20	21
101	HIBEK16	209627 02/12/98	Other	259	1193	69	1135	242	242	561	1	24	25	108
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	112	1313	128	1313	271	271	414	1	30	31	51
102	HE2BG03	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	260	1262	26	1262	35	35	562	1	35	36	50
103	HEBDJ82	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	113	1654	553	1654	709	709	415	1			32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	114	1171	540	1171	337	337	416	1	30	31	163
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	261	1179	626	1161	335	335	563	1	30	31	253
104	HCUBC79	97901 02/26/97 209047 05/15/97	ZAP Express	262	1162	629	1131	942	942	564	1			18
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	115	842	373	800	100	100	417	1	65	66	174
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	263	735	290	735			565	1			
105	HSVAF07	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	264	783	416	783		413	566	1	33	34	73

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	116	1640	187	1470	581	581	418	1	30	31	50
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	265	1638	301	1405	119	119	567	1	30	31	263
106	HT3AM65	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	266	1455	148	1188	438	438	568	1	24	25	70
107	HE6DK18	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	117	952	418	906	499	499	419	1	28	29	120
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	118	1256	21	1079	301	301	420	1	30	31	159
108	HEBEK93	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	267	1086	25	1050	227	227	569	1	23	24	34

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	119	1143	171	1051	175	175	421	1	50	51	154
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	268	1003	21	1003	115	115	570	1	34	35	104
109	HJPCM10	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	269	1234	174	1015	232	232	571	1	27	28	132
110	HSXBL78	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	120	1782	1	1720	138	138	422	1	32	33	204
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	121	610	18	609	50	50	423	1	30	31	67
111	HOEAW81	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	270	574	1	566	337	337	572	1	27	28	32



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
112	HOEAP41	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	122	526	185	375	143	143	424	1	21	22	25
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	123	2081	1179	1976	48	48	425	1	30	31	299
113	HEAAR60	97901 02/26/97 209047 05/15/97	Uni-ZAP XR	271	1731	889	1626	886	886	573	1	18	19	28
114	HTXGS75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	124	1717	764	1640	76	76	426	1			13
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	125	804	1	804	145	145	427	1	15	16	198
115	HOVBA03	97902 02/26/97 209048 05/15/97	pSport1	272	1320	77	637	280	280	574	1	22	23	40

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	126	431	1	431	73	73	428	1	38	39	47
116	HGBGK76	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	273	515	1	515	43	43	575	1	20	21	30
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	127	3752	3465	3752	748	748	429	1	30	31	370
117	HBMUW78	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	274	2995	2738	2995	2777	2777	576	1	18	19	29
118	HASAS24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	128	1144	669	1144	896	896	430	1			30
119	HSIDN55	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	129	1830	1234	1830	1265	1265	431	1			24

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
120	HGBGZ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	130	1864	1505	1741	1578	1578	432	1	37	38	53
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	131	2041	1	1214	46	46	433	1	35	36	176
121	H6EBJ64	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	275	1990	8	1128	71	71	577	1	16	17	92
122	HOECP43	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	132	2012	853	1986	1127	1127	434	1	22	23	77
123	H2CBV31	97902 02/26/97 209048 05/15/97	pBluescript SK-	133	1669	670	1632	962	962	435	1	25	26	32
124	HPCAD23	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	134	1565	281	1565	274	274	436	1	25	26	30

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
125	HSPAG15	97902 02/26/97 209048 05/15/97	pSport1	135	2007	1101	2007	1124	1124	437	1	39	40	69
126	HELGH31	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	136	1291	1	1180	107	107	438	1			19
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	137	1906	1	1906	184	184	439	1	30	31	43
127	HUSHH48	97902 02/26/97 209048 05/15/97	Lambda ZAP II	276	2436	572	2436	726	726	578	1	30	31	42
128	HLYA095	97902 02/26/97 209048 05/15/97	pSport1	138	1935	1044	1794	1183	1183	440	1	18	19	33
129	HHSCV65	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	139	1446	572	1347	585	585	441	1	25	26	53

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
130	HTTAD57	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	140	1109	639	1109	676	676	442	1	24	25	64
131	HEBGA37	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	141	497	9	497	95	95	443	1			34
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	142	269	1	269	1	1	444	1	30	31	89
132	HEBFU93	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	277	782	408	781		571	579	1	31	32	70
133	HSGSC60	97902 02/26/97 209048 05/15/97	Lambda ZAP II	143	1269	55	1262	55	55	445	1	25	26	350
134	HPMGD24	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	144	1944	97	1871	306	306	446	1	16	17	49

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	145	1021	526	1021	74	74	447	1	30	31	278
135	HPTVC60	97902 02/26/97 209048 05/15/97	pBluescript	278	961	524	961	545	545	580	1	23	24	110
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	146	1285	5	1285	116	116	448	1	30	31	199
136	HSKNE18	97902 02/26/97 209048 05/15/97	pBluescript	279	1228	9	1228	324	324	581	1	26	27	30
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	147	1386	169	1272	165	165	449	1	30	31	258
137	HMWIF35	97902 02/26/97 209048 05/15/97	Uni-Zap XR	280	1327	169	1208	160	160	582	1	23	24	71

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
138	HMWGI25	97902 02/26/97 209048 05/15/97	Uni-Zap XR	148	2098	721	2044	784	784	450	1	18	19	87
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	149	1847	1689	1847	241	241	451	1	33	34	315
139	HSKGF03	97902 02/26/97 209048 05/15/97	pBluescript	281	799	1	799		243	583	1	12	13	47
140	HMSKE75	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	150	1569	113	1517	417	417	452	1	21	22	52
141	HCMSSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	151	1540	538	1540	48	48	453	1	30	31	383
141	HCMSSH30	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	282	2196	270	2196	294	294	584	1	32	33	39

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
142	HTWCB92	97902 02/26/97 209048 05/15/97	pSport1	152	1719	690	1575	6	6	454	1	52	53	186
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	153	863	1	863	195	195	455	1	26	27	163
143	HBMDM46	97902 02/26/97 209048 05/15/97	pBluescript	283	1185	277	1166	621	621	585	1			19
144	HFAMG13	97902 02/26/97 209048 05/15/97	Uni-ZAP XR	154	1101	1	512	40	40	456	1	21	22	46
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	155	2031	669	2031	411	411	457	1	23	24	105
145	HFXHL79	97903 02/26/97 209049 05/15/97	Lambda ZAP II	284	1634	615	1485	878	878	586	1	20	21	23

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	156	1981	1458	1809	1592	1592	458	1	23	24	70
146	HSNAK17	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	285	1795	1458	1749	1562	1562	587	1	33	34	69
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	157	915	45	912	22	22	459	1	22	23	155
147	HCFBC03	97903 02/26/97 209049 05/15/97	pSport1	286	858	46	858	224	224	588	1	30	31	77
147	HSJAP03	209139 07/03/97	Uni-ZAP XR	287	915	1	915	22	22	589	1	22	23	155
148	HSKGO26	97903 02/26/97 209049 05/15/97	pBluescript	158	2117	51	1422	32	32	460	1	23	24	332
149	HCQAV96	97903 02/26/97 209049 05/15/97	Lambda ZAP II	159	2395	1509	2382	1440	1440	461	1			5

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
150	HSNCC16	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	160	2120	1223	2108	1416	1416	462	1			14
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	161	900	482	900	46	46	463	1	30	31	285
151	HTLEF62	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	288	1517	783	1517	1062	1062	590	1			24
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	162	1003	1	1003	288	288	464	1	30	31	80
152	HTLAD94	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	289	3865	217	1195	281	281	591	1	16	17	38
153	HTSFQ12	97903 02/26/97 209049 05/15/97	pBluescript	163	2196	1607	2180	1611	1611	465	1	30	31	47

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	164	1945	271	1840	299	299	466	1	63	64	96
154	HE6FL83	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	290	1910	279	1818	355	355	592	1	39	40	69
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	165	2933	489	2871	258	258	467	1	30	31	399
155	HTXFJ55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	291	3276	486	2838		525	593	1	45	46	308
156	HJPCJ76	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	166	2243	343	2221		341	468	1			1
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	167	1816	1130	1816	284	284	469	1	31	32	273

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
157	HLTED27	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	292	1695	1098	1548	1306	1306	594	1			22
158	HMKBA64	97903 02/26/97 209049 05/15/97	pSport1	168	945	1	787	208	208	470	1	18	19	192
159	HNFIP24	97903 02/26/97 209049 05/15/97	pBluescript	169	902	46	816	19	19	471	1	26	27	234
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	170	1883	798	1869	1001	1001	472	1	45	46	105
160	HCELB21	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	293	1501	438	1501	510	510	595	1			24
161	HAWBA28	97903 02/26/97 209049 05/15/97	pBluescript SK-	171	2100	1642	2100	1722	1722	473	1	23	24	32

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	172	1930	187	1930	65	65	474	1	30	31	571
162	HSAAS44	97903 02/26/97 209049 05/15/97	pBluescript SK-	294	2683	183	2683	431	431	596	1			24
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	173	1509	962	1451	122	122	475	1	30	31	312
163	HAFAL73	97903 02/26/97 209049 05/15/97	pBluescript SK-	295	1454	961	1420	976	976	597	1			1
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	174	3173	2197	2972	51	51	476	1	21	22	329
164	HSAWF26	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	296	828	52	828	305	305	598	1			8

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	175	991	374	970	60	60	477	1	24	25	178
165	HEAAL31	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	297	2416	1387	2413	1473	1473	599	1	18	19	25
166	HFKFX55	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	176	1290	499	1290		688	478	1	25	26	52
167	H2LAO11	97903 02/26/97 209049 05/15/97	pBluescript SK-	177	2290	1	2290	173	173	479	1	22	23	62
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	178	549	1	549	11	11	480	1	21	22	27
168	HPFDZ95	97903 02/26/97 209049 05/15/97	Uni-ZAP XR	298	545	1	545	17	17	600	1	21	22	27

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	179	1509	294	1352	92	92	481	1	30	31	339
169	HPTTU11	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	299	1530	385	1530	562	562	601	1	23	24	61
170	HCFAE79	97904 02/26/97 209050 05/15/97	pSport1	180	1316	985	1250	995	995	482	1	26	27	32
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	181	777	1	777	51	51	483	1	30	31	48
171	HTEDJ34	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	300	997	244	997	300	300	602	1	23	24	29
172	HODCW06	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	182	791	1	791	14	14	484	1	29	30	38

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
173	HFTAR26	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	183	1405	346	1405	575	575	485	1	20	21	61
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	184	1596	75	1596	131	131	486	1	24	25	346
174	H2MBF44	97904 02/26/97 209050 05/15/97	pBluescript SK-	301	2345	75	2345	233	233	603	1	56	57	69
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	185	2293	355	2288	67	67	487	1	30	31	237
175	HE8BI92	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	302	2369	2	1946		60	604	1	9	10	24
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	186	1212	462	1180	257	257	488	1	30	31	200



Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
176	HFTBR48	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	303	1181	424	1149	663	663	605	1	23	24	35
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	187	1605	770	1554	166	166	489	1	30	31	351
177	HE9CM64	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	304	1537	719	1515		787	606	1	43	44	130
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	188	1516	960	1516	8	8	490	1	30	31	265
178	HATAV51	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	305	1493	1	1261	54	54	607	1	18	19	23
179	HAQAF27	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	189	681	287	681		401	491	1			25

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	190	1014	703	1014	360	360	492	1	30	31	159
180	HCEEK08	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	306	577	1	577		175	608	1			6
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	191	2779	2207	2630	1153	1153	493	1	30	31	279
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	307	2860	163	2860	21	21	609	1	30	31	232
181	HAFAU18	97904 02/26/97 209050 05/15/97	pBluescript SK-	308	876	275	876	302	302	610	1	32	33	34
182	HETBY74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	192	1923	30	1923	45	45	494	1	33	34	193

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	193	2346	1160	2286	178	178	495	1	30	31	205
183	HTOAF35	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	309	2025	840	2025	971	971	611	1	18	19	21
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	194	3054	2004	3054	434	434	496	1	11	12	147
184	HCRBX32	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	310	3026	1966	3026		2131	612	1			9
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	195	907	152	907	297	297	497	1	30	31	64
185	HEBGB80	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	311	712	67	712	107	107	613	1	18	19	29

Gene No.	cDNA Clone ID	ATCC Deposit No: Z and Date	Vector	NT SEQ ID NO: X	Total NT Seq.	5' NT of Clone Seq.	3' NT of Clone Seq.	5' NT of Start Codon	5' NT of First AA of Signal Pep.	AA SEQ ID NO: Y	First AA of Sig Pep	Last AA of Sig Pep	Predicted First AA of Secreted Portion	Last AA of ORF
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	196	1290	84	809	225	225	498	1	30	31	94
186	HFAMH74	97904 02/26/97 209050 05/15/97	Uni-ZAP XR	312	1289	785	1289	927	927	614	1	28	29	30

Table 1 summarizes the information corresponding to each "Gene No."

described above. The nucleotide sequence identified as "NT SEQ ID NO:X" was assembled from partially homologous ("overlapping") sequences obtained from the "cDNA clone ID" identified in Table 1 and, in some cases, from additional related DNA clones. The overlapping sequences were assembled into a single contiguous sequence of high redundancy (usually three to five overlapping sequences at each nucleotide position), resulting in a final sequence identified as SEQ ID NO:X.

The cDNA Clone ID was deposited on the date and given the corresponding deposit number listed in "ATCC Deposit No:Z and Date." Some of the deposits contain multiple different clones corresponding to the same gene. "Vector" refers to the type of vector contained in the cDNA Clone ID.

"Total NT Seq." refers to the total number of nucleotides in the contig identified by "Gene No." The deposited clone may contain all or most of these sequences, reflected by the nucleotide position indicated as "5' NT of Clone Seq." and the "3' NT of Clone Seq." of SEQ ID NO:X. The nucleotide position of SEQ ID NO:X of the putative start codon (methionine) is identified as "5' NT of Start Codon." Similarly, the nucleotide position of SEQ ID NO:X of the predicted signal sequence is identified as "5' NT of First AA of Signal Pep."

The translated amino acid sequence, beginning with the methionine, is identified as "AA SEQ ID NO:Y," although other reading frames can also be easily translated using known molecular biology techniques. The polypeptides produced by these alternative open reading frames are specifically contemplated by the present invention.

The first and last amino acid position of SEQ ID NO:Y of the predicted signal peptide is identified as "First AA of Sig Pep" and "Last AA of Sig Pep." The predicted first amino acid position of SEQ ID NO:Y of the secreted portion is identified as "Predicted First AA of Secreted Portion." Finally, the amino acid position of SEQ ID NO:Y of the last amino acid in the open reading frame is identified as "Last AA of ORF."

SEQ ID NO:X and the translated SEQ ID NO:Y are sufficiently accurate and otherwise suitable for a variety of uses well known in the art and described further below. For instance, SEQ ID NO:X is useful for designing nucleic acid hybridization probes that will detect nucleic acid sequences contained in SEQ ID NO:X or the cDNA contained in the deposited clone. These probes will also hybridize to nucleic acid molecules in biological samples, thereby enabling a variety of forensic and diagnostic methods of the invention. Similarly, polypeptides identified from SEQ ID NO:Y may be used to generate antibodies which bind specifically to the secreted proteins encoded by the cDNA clones identified in Table 1.

Nevertheless, DNA sequences generated by sequencing reactions can contain sequencing errors. The errors exist as misidentified nucleotides, or as insertions or deletions of nucleotides in the generated DNA sequence. The erroneously inserted or deleted nucleotides cause frame shifts in the reading frames of the predicted amino acid sequence. In these cases, the predicted amino acid sequence diverges from the actual amino acid sequence, even though the generated DNA sequence may be greater than 99.9% identical to the actual DNA sequence (for example, one base insertion or deletion in an open reading frame of over 1000 bases).

Accordingly, for those applications requiring precision in the nucleotide sequence or the amino acid sequence, the present invention provides not only the generated nucleotide sequence identified as SEQ ID NO:X and the predicted translated amino acid sequence identified as SEQ ID NO:Y, but also a sample of plasmid DNA containing a human cDNA of the invention deposited with the ATCC, as set forth in Table 1. The nucleotide sequence of each deposited clone can readily be determined by sequencing the deposited clone in accordance with known methods. The predicted amino acid sequence can then be verified from such deposits. Moreover, the amino acid sequence of the protein encoded by a particular clone can also be directly determined by peptide sequencing or by expressing the protein in a suitable host cell containing the deposited human cDNA, collecting the protein, and determining its sequence.

The present invention also relates to the genes corresponding to SEQ ID NO:X, SEQ ID NO:Y, or the deposited clone. The corresponding gene can be isolated in accordance with known methods using the sequence information disclosed herein. Such methods include preparing probes or primers from the disclosed sequence and identifying or amplifying the corresponding gene from appropriate sources of genomic material.

Also provided in the present invention are species homologs. Species homologs may be isolated and identified by making suitable probes or primers from the sequences provided herein and screening a suitable nucleic acid source for the desired homologue.

The polypeptides of the invention can be prepared in any suitable manner. Such polypeptides include isolated naturally occurring polypeptides, recombinantly produced polypeptides, synthetically produced polypeptides, or polypeptides produced by a combination of these methods. Means for preparing such polypeptides are well understood in the art.

The polypeptides may be in the form of the secreted protein, including the mature form, or may be a part of a larger protein, such as a fusion protein (see below).

It is often advantageous to include an additional amino acid sequence which contains secretory or leader sequences, pro-sequences, sequences which aid in purification, such as multiple histidine residues, or an additional sequence for stability during recombinant production.

The polypeptides of the present invention are preferably provided in an isolated form, and preferably are substantially purified. A recombinantly produced version of a polypeptide, including the secreted polypeptide, can be substantially purified by the one-step method described in Smith and Johnson, *Gene* 67:31-40 (1988).

Polypeptides of the invention also can be purified from natural or recombinant sources using antibodies of the invention raised against the secreted protein in methods which are well known in the art.

### Signal Sequences

Methods for predicting whether a protein has a signal sequence, as well as the cleavage point for that sequence, are available. For instance, the method of McGeoch, *Virus Res.* 3:271-286 (1985), uses the information from a short N-terminal charged region and a subsequent uncharged region of the complete (uncleaved) protein. The method of von Heinjé, *Nucleic Acids Res.* 14:4683-4690 (1986) uses the information from the residues surrounding the cleavage site, typically residues -13 to +2, where +1 indicates the amino terminus of the secreted protein. The accuracy of predicting the cleavage points of known mammalian secretory proteins for each of these methods is in the range of 75-80%. (von Heinjé, *supra*.) However, the two methods do not always produce the same predicted cleavage point(s) for a given protein.

In the present case, the deduced amino acid sequence of the secreted polypeptide was analyzed by a computer program called SignalP (Henrik Nielsen et al., *Protein Engineering* 10:1-6 (1997)), which predicts the cellular location of a protein based on the amino acid sequence. As part of this computational prediction of localization, the methods of McGeoch and von Heinjé are incorporated. The analysis of the amino acid sequences of the secreted proteins described herein by this program provided the results shown in Table 1.

As one of ordinary skill would appreciate, however, cleavage sites sometimes vary from organism to organism and cannot be predicted with absolute certainty. Accordingly, the present invention provides secreted polypeptides having a sequence shown in SEQ ID NO:Y which have an N-terminus beginning within 5 residues (i.e., + or - 5 residues) of the predicted cleavage point. Similarly, it is also recognized that in some cases, cleavage of the signal sequence from a secreted protein is not entirely

uniform, resulting in more than one secreted species. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

Moreover, the signal sequence identified by the above analysis may not necessarily predict the naturally occurring signal sequence. For example, the naturally occurring signal sequence may be further upstream from the predicted signal sequence. However, it is likely that the predicted signal sequence will be capable of directing the secreted protein to the ER. These polypeptides, and the polynucleotides encoding such polypeptides, are contemplated by the present invention.

### Polynucleotide and Polypeptide Variants

"Variant" refers to a polynucleotide or polypeptide differing from the polynucleotide or polypeptide of the present invention, but retaining essential properties thereof. Generally, variants are overall closely similar, and, in many regions, identical to the polynucleotide or polypeptide of the present invention.

"Identity" per se has an art-recognized meaning and can be calculated using published techniques. (See, e.g.: *COMPUTATIONAL MOLECULAR BIOLOGY*, Lesk, A.M., ed., Oxford University Press, New York, (1988); *BIOCOMPUTING: INFORMATICS AND GENOME PROJECTS*, Smith, D.W., ed., Academic Press, New York, (1993); *COMPUTER ANALYSIS OF SEQUENCE DATA, PART I*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, (1994); *SEQUENCE ANALYSIS IN MOLECULAR BIOLOGY*, von Heinjé, G., Academic Press, (1987); and *SEQUENCE ANALYSIS PRIMER*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, (1991).) While there exists a number of methods to measure identity between two polynucleotide or polypeptide sequences, the term "identity" is well known to skilled artisans. (Carillo, H., and Lipton, D., *SIAM J Applied Math* 48:1073 (1988).) Methods commonly employed to determine identity or similarity between two sequences include, but are not limited to, those disclosed in "Guide to Huge Computers," Martin J. Bishop, ed., Academic Press, San Diego, (1994), and Carillo, H., and Lipton, D., *SIAM J Applied Math* 48:1073 (1988).

Methods for aligning polynucleotides or polypeptides are codified in computer programs, including the GCG program package (Devereux, J., et al., *Nucleic Acids Research* (1984) 12(1):387 (1984)), BLASTP, BLASTN, FASTA (Altschul, S.F. et al., *J. Molec. Biol.* 215:403 (1990), Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, WI 53711 (using the local homology algorithm of Smith and Waterman, *Advances in Applied Mathematics* 2:482-489 (1981)).

When using any of the sequence alignment programs to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters are set so that the percentage of identity is calculated over the full length of the reference polynucleotide and that gaps in identity of up to 5% of the total number of nucleotides in the reference polynucleotide are allowed.

A preferred method for determining the best overall match between a query sequence (a sequence of the present invention) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brulag et al. (Comp. App. Biosci. 6:237-245 (1990)). The term "sequence" includes nucleotide and amino acid sequences. In a sequence alignment the query and subject sequences are either both nucleotide sequences or both amino acid sequences. The result of said global sequence alignment is in percent identity. Preferred parameters used in a FASTDB search of a DNA sequence to calculate percent identity are: Matrix=Unitary, k-tuple=4, Mismatch Penalty=1, Joining Penalty=30, Randomization Group Length=0, and Cutoff Score=1, Gap Penalty=5, Gap Size Penalty 0.05, and Window Size=500 or query sequence length in nucleotide bases, whichever is shorter. Preferred parameters employed to calculate percent identity and similarity of an amino acid alignment are: Matrix=PAM 150, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Gap Penalty=5, Gap Size Penalty=0.05, and Window Size=500 or query sequence length in amino acid residues, whichever is shorter.

As an illustration, a polynucleotide having a nucleotide sequence of at least 95% "identity" to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone, means that the polynucleotide is identical to a sequence contained in SEQ ID NO:X or the cDNA except that the polynucleotide sequence may include up to five point mutations per each 100 nucleotides of the total length (not just within a given 100 nucleotide stretch). In other words, to obtain a polynucleotide having a nucleotide sequence at least 95% identical to SEQ ID NO:X or the deposited clone, up to 5% of the nucleotides in the sequence contained in SEQ ID NO:X or the cDNA can be deleted, inserted, or substituted with other nucleotides. These changes may occur anywhere throughout the polynucleotide.

Further embodiments of the present invention include polynucleotides having at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to a sequence contained in SEQ ID NO:X or the cDNA contained in the deposited clone. Of course, due to the degeneracy of the genetic code, one of ordinary skill in the art will immediately recognize that a large number of the polynucleotides having at least 85%, 90%, 95%, 96%, 97%, 98%, or 99% identity

will encode a polypeptide identical to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

Similarly, by a polypeptide having an amino acid sequence having at least, for example, 95% "identity" to a reference polypeptide, is intended that the amino acid sequence of the polypeptide is identical to the reference polypeptide except that the polypeptide sequence may include up to five amino acid alterations per each 100 amino acids of the total length of the reference polypeptide. In other words, to obtain a polypeptide having an amino acid sequence at least 95% identical to a reference amino acid sequence, up to 5% of the amino acid residues in the reference sequence may be deleted or substituted with another amino acid, or a number of amino acids up to 5% of the total amino acid residues in the reference sequence may be inserted into the reference sequence. These alterations of the reference sequence may occur at the amino or carboxy terminal positions of the reference amino acid sequence or anywhere between those terminal positions, interspersed either individually among residues in the reference sequence or in one or more contiguous groups within the reference sequence.

Further embodiments of the present invention include polypeptides having at least 80% identity, more preferably at least 85% identity, more preferably at least 90% identity, and most preferably at least 95%, 96%, 97%, 98% or 99% identity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone. Preferably, the above polypeptides should exhibit at least one biological activity of the protein.

In a preferred embodiment, polypeptides of the present invention include polypeptides having at least 90% similarity, more preferably at least 95% similarity, and still more preferably at least 96%, 97%, 98%, or 99% similarity to an amino acid sequence contained in SEQ ID NO:Y or the expressed protein produced by the deposited clone.

The variants may contain alterations in the coding regions, non-coding regions, or both. Especially preferred are polynucleotide variants containing alterations which produce silent substitutions, additions, or deletions, but do not alter the properties or activities of the encoded polypeptide. Nucleotide variants produced by silent substitutions due to the degeneracy of the genetic code are preferred. Moreover, variants in which 5-10, 1-5, or 1-2 amino acids are substituted, deleted, or added in any combination are also preferred. Polynucleotide variants can be produced for a variety of reasons, e.g., to optimize codon expression for a particular host (change codons in the human mRNA to those preferred by a bacterial host such as *E. coli*).

Naturally occurring variants are called "allelic variants," and refer to one of several alternate forms of a gene occupying a given locus on a chromosome of an

organism. (Genes II, Lewin, B., ed., John Wiley & Sons, New York (1985).) These allelic variants can vary at either the polynucleotide and/or polypeptide level. Alternatively, non-naturally occurring variants may be produced by mutagenesis techniques or by direct synthesis.

5 Using known methods of protein engineering and recombinant DNA

technology, variants may be generated to improve or alter the characteristics of the polypeptides of the present invention. For instance, one or more amino acids can be deleted from the N-terminus or C-terminus of the secreted protein without substantial loss of biological function. The authors of Ron et al., *J. Biol. Chem.* 268: 2984-2988 (1993), reported variant KGF proteins having heparin binding activity even after deleting 3, 8, or 27 amino-terminal amino acid residues. Similarly, Interferon gamma exhibited up to ten times higher activity after deleting 8-10 amino acid residues from the carboxy terminus of this protein. (Dobeli et al., *J. Biotechnology* 7:199-216 (1988).)

10 Moreover, ample evidence demonstrates that variants often retain a biological activity similar to that of the naturally occurring protein. For example, Gayle and coworkers (*J. Biol. Chem.* 268:22105-22111 (1993)) conducted extensive mutational analysis of human cytokine IL-1a. They used random mutagenesis to generate over 3,500 individual IL-1a mutants that averaged 2.5 amino acid changes per variant over the entire length of the molecule. Multiple mutations were examined at every possible amino acid position. The investigators found that "[m]ost of the molecule could be altered with little effect on either [binding or biological activity]." (See, Abstract.) In fact, only 23 unique amino acid sequences, out of more than 3,500 nucleotide sequences examined, produced a protein that significantly differed in activity from wild-type.

25 Furthermore, even if deleting one or more amino acids from the N-terminus or C-terminus of a polypeptide results in modification or loss of one or more biological functions, other biological activities may still be retained. For example, the ability of a deletion variant to induce and/or to bind antibodies which recognize the secreted form will likely be retained when less than the majority of the residues of the secreted form are removed from the N-terminus or C-terminus. Whether a particular polypeptide lacking N- or C-terminal residues of a protein retains such immunogenic activities can readily be determined by routine methods described herein and otherwise known in the art.

35 Thus, the invention further includes polypeptide variants which show substantial biological activity. Such variants include deletions, insertions, inversions, repeats, and substitutions selected according to general rules known in the art so as have little effect on activity. For example, guidance concerning how to make

phenotypically silent amino acid substitutions is provided in Bowie, J. U. et al., *Science* 247:1306-1310 (1990), wherein the authors indicate that there are two main strategies for studying the tolerance of an amino acid sequence to change.

5 The first strategy exploits the tolerance of amino acid substitutions by natural selection during the process of evolution. By comparing amino acid sequences in different species, conserved amino acids can be identified. These conserved amino acids are likely important for protein function. In contrast, the amino acid positions where substitutions have been tolerated by natural selection indicates that these positions are not critical for protein function. Thus, positions tolerating amino acid substitution could be modified while still maintaining biological activity of the protein.

10 The second strategy uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene to identify regions critical for protein function. For example, site directed mutagenesis or alanine-scanning mutagenesis (introduction of single alanine mutations at every residue in the molecule) can be used. (Cunningham and Wells, *Science* 244:1081-1085 (1989).) The resulting mutant molecules can then be tested for biological activity.

15 As the authors state, these two strategies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which amino acid changes are likely to be permissive at certain amino acid positions in the protein. For example, most buried (within the tertiary structure of the protein) amino acid residues require nonpolar side chains, whereas few features of surface side chains are generally conserved. Moreover, tolerated conservative amino acid substitutions involve replacement of the aliphatic or hydrophobic amino acids Ala, Val, Leu and Ile; replacement of the hydroxyl residues Ser and Thr; replacement of the acidic residues Asp and Glu; replacement of the amide residues Asn and Gln; replacement of the basic residues Lys, Arg, and His; replacement of the aromatic residues Phe, Tyr, and Trp, and replacement of the small-sized amino acids Ala, Ser, Thr, Met, and Gly.

25 Besides conservative amino acid substitution, variants of the present invention include (i) substitutions with one or more of the non-conserved amino acid residues, where the substituted amino acid residues may or may not be one encoded by the genetic code, or (ii) substitution with one or more of amino acid residues having a substituent group, or (iii) fusion of the mature polypeptide with another compound, such as a compound to increase the stability and/or solubility of the polypeptide (for example, polyethylene glycol), or (iv) fusion of the polypeptide with additional amino acids, such as an IgG Fc fusion region peptide, or leader or secretory sequence, or a sequence facilitating purification. Such variant polypeptides are deemed to be within the scope of those skilled in the art from the teachings herein.

For example, polypeptide variants containing amino acid substitutions of charged amino acids with other charged or neutral amino acids may produce proteins with improved characteristics, such as less aggregation. Aggregation of pharmaceutical formulations both reduces activity and increases clearance due to the aggregate's immunogenic activity. (Pinckard et al., Clin. Exp. Immunol. 2:331-340 (1967); Robbins et al., Diabetes 36: 838-845 (1987); Cleland et al., Crit. Rev. Therapeutic Drug Carrier Systems 10:307-377 (1993).)

### **Polynucleotide and Polypeptide Fragments**

In the present invention, a "polynucleotide fragment" refers to a short polynucleotide having a nucleic acid sequence contained in the deposited clone or shown in SEQ ID NO:X. The short nucleotide fragments are preferably at least about 15 nt, and more preferably at least about 20 nt, still more preferably at least about 30 nt, and even more preferably, at least about 40 nt in length. A fragment "at least 20 nt in length," for example, is intended to include 20 or more contiguous bases from the cDNA sequence contained in the deposited clone or the nucleotide sequence shown in SEQ ID NO:X. These nucleotide fragments are useful as diagnostic probes and primers as discussed herein. Of course, larger fragments (e.g., 50, 150, 500, 600, 2000 nucleotides) are preferred.

Moreover, representative examples of polynucleotide fragments of the invention, include, for example, fragments having a sequence from about nucleotide number 1-50, 51-100, 101-150, 151-200, 201-250, 251-300, 301-350, 351-400, 401-450, 451-500, 501-550, 551-600, 651-700, and 701 to the end of SEQ ID NO:X or the cDNA contained in the deposited clone. In this context "about" includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) nucleotides, at either terminus or at both termini. Preferably, these fragments encode a polypeptide which has biological activity.

In the present invention, a "polypeptide fragment" refers to a short amino acid sequence contained in SEQ ID NO:Y or encoded by the cDNA contained in the deposited clone. Protein fragments may be "free-standing," or comprised within a larger polypeptide of which the fragment forms a part or region, most preferably as a single continuous region. Representative examples of polypeptide fragments of the invention, include, for example, fragments from about amino acid number 1-20, 21-40, 41-60, 61-80, 81-100, 102-120, 121-140, 141-160, and 161 to the end of the coding region. Moreover, polypeptide fragments can be about 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, or 150 amino acids in length. In this context "about"

includes the particularly recited ranges, larger or smaller by several (5, 4, 3, 2, or 1) amino acids, at either extreme or at both extremes.

Preferred polypeptide fragments include the secreted protein as well as the mature form. Further preferred polypeptide fragments include the secreted protein or the mature form having a continuous series of deleted residues from the amino or the carboxy terminus, or both. For example, any number of amino acids, ranging from 1-60, can be deleted from the amino terminus of either the secreted polypeptide or the mature form. Similarly, any number of amino acids, ranging from 1-30, can be deleted from the carboxy terminus of the secreted protein or mature form. Furthermore, any combination of the above amino and carboxy terminus deletions are preferred. Similarly, polynucleotide fragments encoding these polypeptide fragments are also preferred.

Also preferred are polypeptide and polynucleotide fragments characterized by structural or functional domains, such as fragments that comprise alpha-helix and alpha-helix forming regions, beta-sheet and beta-sheet-forming regions, turn and turn-forming regions, coil and coil-forming regions, hydrophilic regions, hydrophobic regions, alpha amphipathic regions, beta amphipathic regions, flexible regions, surface-forming regions, substrate binding region, and high antigenic index regions. Polypeptide fragments of SEQ ID NO:Y falling within conserved domains are specifically contemplated by the present invention. Moreover, polynucleotide fragments encoding these domains are also contemplated.

Other preferred fragments are biologically active fragments. Biologically active fragments are those exhibiting activity similar, but not necessarily identical, to an activity of the polypeptide of the present invention. The biological activity of the fragments may include an improved desired activity, or a decreased undesirable activity.

### **Epitopes & Antibodies**

In the present invention, "epitopes" refer to polypeptide fragments having antigenic or immunogenic activity in an animal, especially in a human. A preferred embodiment of the present invention relates to a polypeptide fragment comprising an epitope, as well as the polynucleotide encoding this fragment. A region of a protein molecule to which an antibody can bind is defined as an "antigenic epitope." In contrast, an "immunogenic epitope" is defined as a part of a protein that elicits an antibody response. (See, for instance, Geysen et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1983).)

Fragments which function as epitopes may be produced by any conventional means. (See, e.g., Houghten, R. A., *Proc. Natl. Acad. Sci. USA* 82:5131-5135 (1985) further described in U.S. Patent No. 4,631,211.)

In the present invention, antigenic epitopes preferably contain a sequence of at least seven, more preferably at least nine, and most preferably between about 15 to about 30 amino acids. Antigenic epitopes are useful to raise antibodies, including monoclonal antibodies, that specifically bind the epitope. (See, for instance, Wilson et al., *Cell* 37:767-778 (1984); Sutcliffe, J. G. et al., *Science* 219:660-666 (1983).)

Similarly, immunogenic epitopes can be used to induce antibodies according to methods well known in the art. (See, for instance, Sutcliffe et al., *supra*; Wilson et al., *supra*; Chow, M. et al., *Proc. Natl. Acad. Sci. USA* 82:910-914; and Bittle, F. J. et al., *J. Gen. Virol.* 66:2347-2354 (1985).) A preferred immunogenic epitope includes the secreted protein. The immunogenic epitopes may be presented together with a carrier protein, such as an albumin, to an animal system (such as rabbit or mouse) or, if it is long enough (at least about 25 amino acids), without a carrier. However, immunogenic epitopes comprising as few as 8 to 10 amino acids have been shown to be sufficient to raise antibodies capable of binding to, at the very least, linear epitopes in a denatured polypeptide (e.g., in Western blotting.)

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')<sub>2</sub> fragments) which are capable of specifically binding to protein. Fab and F(ab')<sub>2</sub> fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding than an intact antibody. (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983).) Thus, these fragments are preferred, as well as the products of a FAB or other immunoglobulin expression library.

Moreover, antibodies of the present invention include chimeric, single chain, and humanized antibodies.

### Fusion Proteins

Any polypeptide of the present invention can be used to generate fusion proteins. For example, the polypeptide of the present invention, when fused to a second protein, can be used as an antigenic tag. Antibodies raised against the polypeptide of the present invention can be used to indirectly detect the second protein by binding to the polypeptide. Moreover, because secreted proteins target cellular locations based on trafficking signals, the polypeptides of the present invention can be used as targeting molecules once fused to other proteins.

Examples of domains that can be fused to polypeptides of the present invention include not only heterologous signal sequences, but also other heterologous functional regions. The fusion does not necessarily need to be direct, but may occur through linker sequences.

Moreover, fusion proteins may also be engineered to improve characteristics of the polypeptide of the present invention. For instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence during purification from the host cell or subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification. Such regions may be removed prior to final preparation of the polypeptide. The addition of peptide moieties to facilitate handling of polypeptides are familiar and routine techniques in the art.

Moreover, polypeptides of the present invention, including fragments, and specifically epitopes, can be combined with parts of the constant domain of immunoglobulins (IgG), resulting in chimeric polypeptides. These fusion proteins facilitate purification and show an increased half-life in vivo. One reported example describes chimeric proteins consisting of the first two domains of the human CD4-polypeptide and various domains of the constant regions of the heavy or light chains of mammalian immunoglobulins. (EP A 394,827; Trautnecker et al., *Nature* 331:84-86 (1988).) Fusion proteins having disulfide-linked dimeric structures (due to the IgG) can also be more efficient in binding and neutralizing other molecules, than the monomeric secreted protein or protein fragment alone. (Fountoulakis et al., *J. Biochem.* 270:3958-3964 (1995).)

Similarly, EP-A-O 464 533 (Canadian counterpart 2045869) discloses fusion proteins comprising various portions of constant region of immunoglobulin molecules together with another human protein or part thereof. In many cases, the Fc part in a fusion protein is beneficial in therapy and diagnosis, and thus can result in, for example, improved pharmacokinetic properties. (EP-A 0232 262.) Alternatively, deleting the Fc part after the fusion protein has been expressed, detected, and purified, would be desired. For example, the Fc portion may hinder therapy and diagnosis if the fusion protein is used as an antigen for immunizations. In drug discovery, for example, human proteins, such as hIL-5, have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists of hIL-5. (See, D. Bennett et al., *J. Molecular Recognition* 8:52-58 (1995); K. Johanson et al., *J. Biol. Chem.* 270:9459-9471 (1995).)

Moreover, the polypeptides of the present invention can be fused to marker sequences, such as a peptide which facilitates purification of the fused polypeptide. In



preferred embodiments, the marker amino acid sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (QIAGEN, Inc., 9259 Eton Avenue, Chatsworth, CA, 91311), among others, many of which are commercially available. As described in Gentz et al., Proc. Natl. Acad. Sci. USA 86:821-824 (1989), for instance, hexa-histidine provides for convenient purification of the fusion protein.

Another peptide tag useful for purification, the "HA" tag, corresponds to an epitope derived from the influenza hemagglutinin protein. (Wilson et al., Cell 37:767 (1984).) Thus, any of these above fusions can be engineered using the polynucleotides or the polypeptides of the claimed invention.

### Vectors, Host Cells, and Protein Production

The present invention also relates to vectors containing the polynucleotide of the present invention, host cells, and the production of polypeptides by recombinant techniques. The vector may be, for example, a phage, plasmid, viral, or retroviral vector. Retroviral vectors may be replication competent or replication defective. In the latter case, viral propagation generally will occur only in complementing host cells.

The polynucleotides may be joined to a vector containing a selectable marker for propagation in a host. Generally, a plasmid vector is introduced in a precipitate, such as a calcium phosphate precipitate, or in a complex with a charged lipid. If the vector is a virus, it may be packaged *in vitro* using an appropriate packaging cell line and then transduced into host cells.

The polynucleotide insert should be operatively linked to an appropriate promoter, such as the phage lambda P<sub>L</sub> promoter, the E. coli lac, trp, phoA and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled artisan. The expression constructs will further contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the transcripts expressed by the constructs will preferably include a translation initiating codon at the beginning and a termination codon (UAA, UGA or UAG) appropriately positioned at the end of the polypeptide to be translated.

As indicated, the expression vectors will preferably include at least one selectable marker. Such markers include dihydrofolate reductase, G418 or neomycin resistance for eukaryotic cell culture and tetracycline, kanamycin or ampicillin resistance genes for culturing in E. coli and other bacteria. Representative examples of appropriate hosts include, but are not limited to, bacterial cells, such as E. coli, Streptomyces and Salmonella typhimurium cells; fungal cells, such as yeast cells; insect cells such as Drosophila S2 and Spodoptera Sf9 cells; animal cells such as CHO, COS,

293, and Bowes melanoma cells; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria include pQE70, pQE60 and pQE-9, available from QIAGEN, Inc.; pBluescript vectors, Phagescript vectors, pNH8A, pNH16a, pNH18A, pNH46A, available from Stratagene Cloning Systems, Inc.; and ptc99a, pKK223-3, pKK233-3, pDR340, pRIT5 available from Pharmacia Biotech, Inc. Among preferred eukaryotic vectors are pWLNEO, pSV2CAT, pOG44, pXT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Introduction of the construct into the host cell can be effected by calcium phosphate transfection, DEAE-dextran mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, or other methods. Such methods are described in many standard laboratory manuals, such as Davis et al., Basic Methods In Molecular Biology (1986). It is specifically contemplated that the polypeptides of the present invention may in fact be expressed by a host cell lacking a recombinant vector.

A polypeptide of this invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, affinity phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

Polypeptides of the present invention, and preferably the secreted form, can also be recovered from: products purified from natural sources, including bodily fluids, tissues and cells, whether directly isolated or cultured; products of chemical synthetic procedures; and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect, and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

Thus, it is well known in the art that the N-terminal methionine encoded by the translation initiation codon generally is removed with high efficiency from any protein after translation in all eukaryotic cells. While the N-terminal methionine on most proteins also is efficiently removed in most prokaryotes, for some proteins, this prokaryotic removal process is inefficient, depending on the nature of the amino acid to which the N-terminal methionine is covalently linked.

### Uses of the Polynucleotides

Each of the polynucleotides identified herein can be used in numerous ways as reagents. The following description should be considered exemplary and utilizes known techniques.

The polynucleotides of the present invention are useful for chromosome identification. There exists an ongoing need to identify new chromosome markers, since few chromosome marking reagents, based on actual sequence data (repeat polymorphisms), are presently available. Each polynucleotide of the present invention can be used as a chromosome marker.

Briefly, sequences can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp) from the sequences shown in SEQ ID NO:X. Primers can be selected using computer analysis so that primers do not span more than one predicted exon in the genomic DNA. These primers are then used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the SEQ ID NO:X will yield an amplified fragment.

Similarly, somatic hybrids provide a rapid method of PCR mapping the polynucleotides to particular chromosomes. Three or more clones can be assigned per day using a single thermal cycler. Moreover, sublocalization of the polynucleotides can be achieved with panels of specific chromosome fragments. Other gene mapping strategies that can be used include *in situ* hybridization, prescreening with labeled flow-sorted chromosomes, and prescreening by hybridization to construct chromosome specific-cDNA libraries.

Precise chromosomal location of the polynucleotides can also be achieved using fluorescence *in situ* hybridization (FISH) of a metaphase chromosomal spread. This technique uses polynucleotides as short as 500 or 600 bases; however, polynucleotides 2,000-4,000 bp are preferred. For a review of this technique, see Verma et al., "Human Chromosomes: a Manual of Basic Techniques," Pergamon Press, New York (1988).

For chromosome mapping, the polynucleotides can be used individually (to mark a single chromosome or a single site on that chromosome) or in panels (for marking multiple sites and/or multiple chromosomes). Preferred polynucleotides correspond to the noncoding regions of the cDNAs because the coding sequences are more likely conserved within gene families, thus increasing the chance of cross hybridization during chromosomal mapping.

Once a polynucleotide has been mapped to a precise chromosomal location, the physical position of the polynucleotide can be used in linkage analysis. Linkage

analysis establishes coinherence between a chromosomal location and presentation of a particular disease. (Disease mapping data are found, for example, in V. McKusick, Mendelian Inheritance in Man (available on line through Johns Hopkins University Welch Medical Library).) Assuming 1 megabase mapping resolution and one gene per 20 kb, a cDNA precisely localized to a chromosomal region associated with the disease could be one of 50-500 potential causative genes.

Thus, once coinherence is established, differences in the polynucleotide and the corresponding gene between affected and unaffected individuals can be examined.

First, visible structural alterations in the chromosomes, such as deletions or translocations, are examined in chromosome spreads or by PCR. If no structural alterations exist, the presence of point mutations are ascertained. Mutations observed in some or all affected individuals, but not in normal individuals, indicates that the mutation may cause the disease. However, complete sequencing of the polypeptide and the corresponding gene from several normal individuals is required to distinguish the mutation from a polymorphism. If a new polymorphism is identified, this polymorphic polypeptide can be used for further linkage analysis.

Furthermore, increased or decreased expression of the gene in affected individuals as compared to unaffected individuals can be assessed using polynucleotides of the present invention. Any of these alterations (altered expression, chromosomal rearrangement, or mutation) can be used as a diagnostic or prognostic marker.

In addition to the foregoing, a polynucleotide can be used to control gene expression through triple helix formation or antisense DNA or RNA. Both methods rely on binding of the polynucleotide to DNA or RNA. For these techniques, preferred polynucleotides are usually 20 to 40 bases in length and complementary to either the region of the gene involved in transcription (triple helix - see Lee et al., Nucl. Acids Res. 6:3073 (1979); Cooney et al., Science 241:456 (1988); and Dervan et al., Science 251:1360 (1991)) or to the mRNA itself (antisense - Okano, J. Neurochem. 56:560 (1991); Oligodeoxy-nucleotides as Antisense Inhibitors of Gene Expression, CRC Press, Boca Raton, FL (1988).) Triple helix formation optimally results in a shut-off of RNA transcription from DNA, while antisense RNA hybridization blocks translation of an mRNA molecule into polypeptide. Both techniques are effective in model systems, and the information disclosed herein can be used to design antisense or triple helix polynucleotides in an effort to treat disease.

Polynucleotides of the present invention are also useful in gene therapy. One goal of gene therapy is to insert a normal gene into an organism having a defective gene, in an effort to correct the genetic defect. The polynucleotides disclosed in the

present invention offer a means of targeting such genetic defects in a highly accurate manner. Another goal is to insert a new gene that was not present in the host genome, thereby producing a new trait in the host cell.

The polynucleotides are also useful for identifying individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identifying personnel. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The polynucleotides of the present invention can be used as additional DNA markers for RFLP.

The polynucleotides of the present invention can also be used as an alternative to RFLP, by determining the actual base-by-base DNA sequence of selected portions of an individual's genome. These sequences can be used to prepare PCR primers for amplifying and isolating such selected DNA, which can then be sequenced. Using this technique, individuals can be identified because each individual will have a unique set of DNA sequences. Once an unique ID database is established for an individual, positive identification of that individual, living or dead, can be made from extremely small tissue samples.

Forensic biology also benefits from using DNA-based identification techniques as disclosed herein. DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, semen, etc., can be amplified using PCR. In one prior art technique, gene sequences amplified from polymorphic loci, such as DQa class II HLA gene, are used in forensic biology to identify individuals. (Erich, H., PCR Technology, Freeman and Co. (1992).) Once these specific polymorphic loci are amplified, they are digested with one or more restriction enzymes, yielding an identifying set of bands on a Southern blot probed with DNA corresponding to the DQa class II HLA gene. Similarly, polynucleotides of the present invention can be used as polymorphic markers for forensic purposes.

There is also a need for reagents capable of identifying the source of a particular tissue. Such need arises, for example, in forensics when presented with tissue of unknown origin. Appropriate reagents can comprise, for example, DNA probes or primers specific to particular tissue prepared from the sequences of the present invention. Panels of such reagents can identify tissue by species and/or by organ type. In a similar fashion, these reagents can be used to screen tissue cultures for contamination.

In the very least, the polynucleotides of the present invention can be used as molecular weight markers on Southern gels, as diagnostic probes for the presence of a specific mRNA in a particular cell type, as a probe to "subtract-out" known sequences in the process of discovering novel polynucleotides, for selecting and making oligomers for attachment to a "gene chip" or other support, to raise anti-DNA antibodies using DNA immunization techniques, and as an antigen to elicit an immune response.

### Uses of the Polypeptides

Each of the polypeptides identified herein can be used in numerous ways. The following description should be considered exemplary and utilizes known techniques.

A polypeptide of the present invention can be used to assay protein levels in a biological sample using antibody-based techniques. For example, protein expression in tissues can be studied with classical immunohistological methods. (Jalkanen, M., et al., J. Cell. Biol. 101:976-985 (1985); Jalkanen, M., et al., J. Cell. Biol. 105:3087-3096 (1987).) Other antibody-based methods useful for detecting protein gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). Suitable antibody assay labels are known in the art and include enzyme labels, such as, glucose oxidase, and radioisotopes, such as iodine (<sup>125</sup>I, <sup>121</sup>I), carbon (<sup>14</sup>C), sulfur (<sup>35</sup>S), tritium (<sup>3</sup>H), indium (<sup>112</sup>In), and technetium (<sup>99m</sup>Tc), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

In addition to assaying secreted protein levels in a biological sample, proteins can also be detected in vivo by imaging. Antibody labels or markers for in vivo imaging of protein include those detectable by X-radiography, NMR or ESR. For X-radiography, suitable labels include radioisotopes such as barium or cesium, which emit detectable radiation but are not overtly harmful to the subject. Suitable markers for NMR and ESR include those with a detectable characteristic spin, such as deuterium, which may be incorporated into the antibody by labeling of nutrients for the relevant hydridoma.

A protein-specific antibody or antibody fragment which has been labeled with an appropriate detectable imaging moiety, such as a radioisotope (for example, <sup>131</sup>I, <sup>112</sup>In, <sup>99m</sup>Tc), a radio-opaque substance, or a material detectable by nuclear magnetic resonance, is introduced (for example, parenterally, subcutaneously, or intraperitoneally) into the mammal. It will be understood in the art that the size of the subject and the imaging system used will determine the quantity of imaging moiety needed to produce diagnostic images. In the case of a radioisotope moiety, for a human subject, the quantity of radioactivity injected will normally range from about 5 to 20

millicuries of  $^{99m}\text{Tc}$ . The labeled antibody or antibody fragment will then preferentially accumulate at the location of cells which contain the specific protein. In vivo tumor imaging is described in S.W. Burchiel et al., "Immunopharmacokinetics of Radiolabeled Antibodies and Their Fragments." (Chapter 13 in Tumor Imaging: The Radiochemical Detection of Cancer, S.W. Burchiel and B. A. Rhodes, eds., Masson Publishing Inc. (1982).)

Thus, the invention provides a diagnostic method of a disorder, which involves (a) assaying the expression of a polypeptide of the present invention in cells or body fluid of an individual; (b) comparing the level of gene expression with a standard gene expression level, whereby an increase or decrease in the assayed polypeptide gene expression level compared to the standard expression level is indicative of a disorder.

Moreover, polypeptides of the present invention can be used to treat disease. For example, patients can be administered a polypeptide of the present invention in an effort to replace absent or decreased levels of the polypeptide (e.g., insulin), to supplement absent or decreased levels of a different polypeptide (e.g., hemoglobin S for hemoglobin B), to inhibit the activity of a polypeptide (e.g., an oncogene), to activate the activity of a polypeptide (e.g., by binding to a receptor), to reduce the activity of a membrane bound receptor by competing with it for free ligand (e.g., soluble TNF receptors used in reducing inflammation), or to bring about a desired response (e.g., blood vessel growth).

Similarly, antibodies directed to a polypeptide of the present invention can also be used to treat disease. For example, administration of an antibody directed to a polypeptide of the present invention can bind and reduce overproduction of the polypeptide. Similarly, administration of an antibody can activate the polypeptide, such as by binding to a polypeptide bound to a membrane (receptor).

At the very least, the polypeptides of the present invention could be used as molecular weight markers on SDS-PAGE gels or on molecular sieve gel filtration columns using methods well known to those of skill in the art. Polypeptides can also be used to raise antibodies, which in turn are used to measure protein expression from a recombinant cell, as a way of assessing transformation of the host cell. Moreover, the polypeptides of the present invention can be used to test the following biological activities.

### Biological Activities

The polynucleotides and polypeptides of the present invention can be used in assays to test for one or more biological activities. If these polynucleotides and polypeptides do exhibit activity in a particular assay, it is likely that these molecules

may be involved in the diseases associated with the biological activity. Thus, the polynucleotides and polypeptides could be used to treat the associated disease.

### Immune Activity

A polypeptide or polynucleotide of the present invention may be useful in treating deficiencies or disorders of the immune system, by activating or inhibiting the proliferation, differentiation, or mobilization (chemotaxis) of immune cells. Immune cells develop through a process called hematopoiesis, producing myeloid (platelets, red blood cells, neutrophils, and macrophages) and lymphoid (B and T lymphocytes) cells from pluripotent stem cells. The etiology of these immune deficiencies or disorders may be genetic, somatic, such as cancer or some autoimmune disorders, acquired (e.g., by chemotherapy or toxins), or infectious. Moreover, a polynucleotide or polypeptide of the present invention can be used as a marker or detector of a particular immune system disease or disorder.

A polynucleotide or polypeptide of the present invention may be useful in treating or detecting deficiencies or disorders of hematopoietic cells. A polypeptide or polynucleotide of the present invention could be used to increase differentiation and proliferation of hematopoietic cells, including the pluripotent stem cells, in an effort to treat those disorders associated with a decrease in certain (or many) types hematopoietic cells. Examples of immunologic deficiency syndromes include, but are not limited to: blood protein disorders (e.g. agammaglobulinemia, dysgammaglobulinemia), ataxia telangiectasia, common variable immunodeficiency, DiGeorge Syndrome, HIV infection, HTLV-BLV infection, leukocyte adhesion deficiency syndrome, lymphopenia, phagocyte bactericidal dysfunction, severe combined immunodeficiency (SCIDs), Wiskott-Aldrich Disorder, anemia, thrombocytopenia, or hemoglobinuria.

Moreover, a polypeptide or polynucleotide of the present invention could also be used to modulate hemostatic (the stopping of bleeding) or thrombolytic activity (clot formation). For example, by increasing hemostatic or thrombolytic activity, a polynucleotide or polypeptide of the present invention could be used to treat blood coagulation disorders (e.g., afibrinogenemia, factor deficiencies), blood platelet disorders (e.g. thrombocytopenia), or wounds resulting from trauma, surgery, or other causes. Alternatively, a polynucleotide or polypeptide of the present invention that can decrease hemostatic or thrombolytic activity could be used to inhibit or dissolve clotting. These molecules could be important in the treatment of heart attacks (infarction), strokes, or scarring.

A polynucleotide or polypeptide of the present invention may also be useful in treating or detecting autoimmune disorders. Many autoimmune disorders result from

inappropriate recognition of self as foreign material by immune cells. This inappropriate recognition results in an immune response leading to the destruction of the host tissue. Therefore, the administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing autoimmune disorders.

Examples of autoimmune disorders that can be treated or detected by the present invention include, but are not limited to: Addison's Disease, hemolytic anemia, anaphospholipid syndrome, rheumatoid arthritis, dermatitis, allergic encephalomyelitis, glomerulonephritis, Goodpasture's Syndrome, Graves' Disease, Multiple Sclerosis, Myasthenia Gravis, Neuritis, Ophthalmia, Bullous Pemphigoid, Pemphigus, Polyendocrinopathies, Purpura, Reiter's Disease, Stiff-Man Syndrome, Autoimmune Thyroiditis, Systemic Lupus Erythematosus, Autoimmune Pulmonary Inflammation, Guillain-Barre Syndrome, insulin dependent diabetes mellitus, and autoimmune inflammatory eye disease.

Similarly, allergic reactions and conditions, such as asthma (particularly allergic asthma) or other respiratory problems, may also be treated by a polypeptide or polynucleotide of the present invention. Moreover, these molecules can be used to treat anaphylaxis, hypersensitivity to an antigenic molecule, or blood group incompatibility.

A polynucleotide or polypeptide of the present invention may also be used to treat and/or prevent organ rejection or graft-versus-host disease (GVHD). Organ rejection occurs by host immune cell destruction of the transplanted tissue through an immune response. Similarly, an immune response is also involved in GVHD, but, in this case, the foreign transplanted immune cells destroy the host tissues. The administration of a polypeptide or polynucleotide of the present invention that inhibits an immune response, particularly the proliferation, differentiation, or chemotaxis of T-cells, may be an effective therapy in preventing organ rejection or GVHD.

Similarly, a polypeptide or polynucleotide of the present invention may also be used to modulate inflammation. For example, the polypeptide or polynucleotide may inhibit the proliferation and differentiation of cells involved in an inflammatory response. These molecules can be used to treat inflammatory conditions, both chronic and acute conditions, including inflammation associated with infection (e.g., septic shock, sepsis, or systemic inflammatory response syndrome (SIRS)), ischemia-reperfusion injury, endotoxin lethality, arthritis, complement-mediated hyperacute rejection, nephritis, cytokine or chemokine induced lung injury, inflammatory bowel disease, Crohn's disease, or resulting from over production of cytokines (e.g., TNF or IL-1).

### **Hyperproliferative Disorders**

A polypeptide or polynucleotide can be used to treat or detect hyperproliferative disorders, including neoplasms. A polypeptide or polynucleotide of the present invention may inhibit the proliferation of the disorder through direct or indirect interactions. Alternatively, a polypeptide or polynucleotide of the present invention may proliferate other cells which can inhibit the hyperproliferative disorder.

For example, by increasing an immune response, particularly increasing antigenic qualities of the hyperproliferative disorder or by proliferating, differentiating, or mobilizing T-cells, hyperproliferative disorders can be treated. This immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, decreasing an immune response may also be a method of treating hyperproliferative disorders, such as a chemotherapeutic agent.

Examples of hyperproliferative disorders that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but are not limited to neoplasms located in the: abdomen, bone, breast, digestive system, liver, pancreas, peritoneum, endocrine glands (adrenal, parathyroid, pituitary, testicles, ovary, thymus, thyroid), eye, head and neck, nervous (central and peripheral), lymphatic system, pelvic, skin, soft tissue, spleen, thoracic, and urogenital.

Similarly, other hyperproliferative disorders can also be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of such hyperproliferative disorders include, but are not limited to: hypergammaglobulinemia, lymphoproliferative disorders, paraproteinemias, purpura, sarcoidosis, Sezary Syndrome, Waldenstrom's Macroglobulinemia, Gaucher's Disease, histiocytosis, and any other hyperproliferative disease, besides neoplasia, located in an organ system listed above.

### **Infectious Disease**

A polypeptide or polynucleotide of the present invention can be used to treat or detect infectious agents. For example, by increasing the immune response, particularly increasing the proliferation and differentiation of B and/or T cells, infectious diseases may be treated. The immune response may be increased by either enhancing an existing immune response, or by initiating a new immune response. Alternatively, the polypeptide or polynucleotide of the present invention may also directly inhibit the infectious agent, without necessarily eliciting an immune response.

Viruses are one example of an infectious agent that can cause disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention. Examples of viruses, include, but are not limited to the following DNA and RNA viral families: Arbovirus, Adenoviridae, Arenaviridae, Arterivirus, Birnaviridae, Bunyaviridae, Caliciviridae, Circoviridae, Coronaviridae, Flaviviridae, Hepadnaviridae (Hepatitis), Herpesviridae (such as, Cytomegalovirus, Herpes Simplex, Herpes Zoster), Mononegavirus (e.g., Paramyxoviridae, Morbillivirus, Rhabdoviridae), Orthomyxoviridae (e.g., Influenza), Papovaviridae, Parvoviridae, Picornaviridae, Poxviridae (such as Smallpox or Vaccinia), Reoviridae (e.g., Rotavirus), Retroviridae (HTLV-I, HTLV-II, Lentivirus), and Togaviridae (e.g., Rubivirus). Viruses falling within these families can cause a variety of diseases or symptoms, including, but not limited to: arthritis, bronchiolitis, encephalitis, eye infections (e.g., conjunctivitis, keratitis), chronic fatigue syndrome, hepatitis (A, B, C, E, Chronic Active, Delta), meningitis, opportunistic infections (e.g., AIDS), pneumonia, Burkitt's Lymphoma, chickenpox, hemorrhagic fever, Measles, Mumps, Parainfluenza, Rabies, the common cold, Polio, leukemia, Rubella, sexually transmitted diseases, skin diseases (e.g., Kaposi's, warts), and viremia. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Similarly, bacterial or fungal agents that can cause disease or symptoms and that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following Gram-Negative and Gram-positive bacterial families and fungi: Actinomycetales (e.g., *Corynebacterium*, *Mycobacterium*, *Nocardia*), *Aspergillus*, *Bacillaceae* (e.g., *Anthrax*, *Clostridium*), *Bacteroidaceae*, *Blastomycosis*, *Bordetella*, *Borrelia*, *Brucellosis*, *Candidiasis*, *Campylobacter*, *Coccidioidomycosis*, *Cryptococcosis*, *Dermatocycoses*, *Enterobacteriaceae* (*Klebsiella*, *Salmonella*, *Serratia*, *Yersinia*), *Erysipelothrix*, *Helicobacter*, *Legionellosis*, *Leptospirosis*, *Listeria*, *Mycoplasmales*, *Neisseriaceae* (e.g., *Acinetobacter*, *Gonorrhea*, *Menigococcal*), *Pasteurellaceae* (*Actinobacillus*, *Haemophilus*, *Pasteurella*), *Pseudomonas*, *Rickettsiaceae*, *Chlamydiaceae*, *Syphilis*, and *Staphylococcal*. These bacterial or fungal families can cause the following diseases or symptoms, including, but not limited to: bacteremia, endocarditis, eye infections (conjunctivitis, tuberculosis, uveitis), gingivitis, opportunistic infections (e.g., AIDS related infections), paronychia, prosthesis-related infections, Reiter's Disease, respiratory tract infections, such as Whooping Cough or Empyema, sepsis, Lyme Disease, Cat-Scratch Disease, Dysentery, Paratyphoid Fever, food poisoning, Typhoid, pneumonia, Gonorrhea, meningitis, Chlamydia, Syphilis, Diphtheria,

Leprosy, Paratuberculosis, Tuberculosis, Lupus, Botulism, gangrene, tetanus, impetigo, Rheumatic Fever, Scarlet Fever, sexually transmitted diseases, skin diseases (e.g., cellulitis, dermatocycoses), toxemia, urinary tract infections, wound infections. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Moreover, parasitic agents causing disease or symptoms that can be treated or detected by a polynucleotide or polypeptide of the present invention include, but not limited to, the following families: Amebiasis, Babesiosis, Coccidiosis, Cryptosporidiosis, Dientamoebiasis, Dourine, Ectoparasitic, Giardiasis, Helminthiasis, Leishmaniasis, Theileriasis, Toxoplasmosis, Trypanosomiasis, and Trichomonas. These parasites can cause a variety of diseases or symptoms, including, but not limited to: Scabies, Trombiculiasis, eye infections, intestinal disease (e.g., dysentery, giardiasis), liver disease, lung disease, opportunistic infections (e.g., AIDS related), Malaria, pregnancy complications, and toxoplasmosis. A polypeptide or polynucleotide of the present invention can be used to treat or detect any of these symptoms or diseases.

Preferably, treatment using a polypeptide or polynucleotide of the present invention could either be by administering an effective amount of a polypeptide to the patient, or by removing cells from the patient, supplying the cells with a polynucleotide of the present invention, and returning the engineered cells to the patient (ex vivo therapy). Moreover, the polypeptide or polynucleotide of the present invention can be used as an antigen in a vaccine to raise an immune response against infectious disease.

### Regeneration

A polynucleotide or polypeptide of the present invention can be used to differentiate, proliferate, and attract cells, leading to the regeneration of tissues. (See, Science 276:59-87 (1997).) The regeneration of tissues could be used to repair, replace, or protect tissue damaged by congenital defects, trauma (wounds, burns, incisions, or ulcers), age, disease (e.g. osteoporosis, osteoarthritis, periodontal disease, liver failure), surgery, including cosmetic plastic surgery, fibrosis, reperfusion injury, or systemic cytokine damage.

Tissues that could be regenerated using the present invention include organs (e.g., pancreas, liver, intestine, kidney, skin, endothelium), muscle (smooth, skeletal or cardiac), vascular (including vascular endothelium), nervous, hematopoietic, and skeletal (bone, cartilage, tendon, and ligament) tissue. Preferably, regeneration occurs without or decreased scarring. Regeneration also may include angiogenesis.

Moreover, a polynucleotide or polypeptide of the present invention may increase regeneration of tissues difficult to heal. For example, increased tendon/ligament regeneration would quicken recovery time after damage. A polynucleotide or polypeptide of the present invention could also be used prophylactically in an effort to avoid damage. Specific diseases that could be treated include of tendinitis, carpal tunnel syndrome, and other tendon or ligament defects. A further example of tissue regeneration of non-healing wounds includes pressure ulcers, ulcers associated with vascular insufficiency, surgical, and traumatic wounds.

Similarly, nerve and brain tissue could also be regenerated by using a polynucleotide or polypeptide of the present invention to proliferate and differentiate nerve cells. Diseases that could be treated using this method include central and peripheral nervous system diseases, neuropathies, or mechanical and traumatic disorders (e.g., spinal cord disorders, head trauma, cerebrovascular disease, and stroke). Specifically, diseases associated with peripheral nerve injuries, peripheral neuropathy (e.g., resulting from chemotherapy or other medical therapies), localized neuropathies, and central nervous system diseases (e.g., Alzheimer's disease, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager syndrome), could all be treated using the polynucleotide or polypeptide of the present invention.

### Chemotaxis

A polynucleotide or polypeptide of the present invention may have chemotaxis activity. A chemotactic molecule attracts or mobilizes cells (e.g., monocytes, fibroblasts, neutrophils, T-cells, mast cells, eosinophils, epithelial and/or endothelial cells) to a particular site in the body, such as inflammation, infection, or site of hyperproliferation. The mobilized cells can then fight off and/or heal the particular trauma or abnormality.

A polynucleotide or polypeptide of the present invention may increase chemotactic activity of particular cells. These chemotactic molecules can then be used to treat inflammation, infection, hyperproliferative disorders, or any immune system disorder by increasing the number of cells targeted to a particular location in the body. For example, chemotactic molecules can be used to treat wounds and other trauma to tissues by attracting immune cells to the injured location. Chemotactic molecules of the present invention can also attract fibroblasts, which can be used to treat wounds.

It is also contemplated that a polynucleotide or polypeptide of the present invention may inhibit chemotactic activity. These molecules could also be used to treat

disorders. Thus, a polynucleotide or polypeptide of the present invention could be used as an inhibitor of chemotaxis.

### Binding Activity

A polypeptide of the present invention may be used to screen for molecules that bind to the polypeptide or for molecules to which the polypeptide binds. The binding of the polypeptide and the molecule may activate (agonist), increase, inhibit (antagonist), or decrease activity of the polypeptide or the molecule bound. Examples of such molecules include antibodies, oligonucleotides, proteins (e.g., receptors), or small molecules.

Preferably, the molecule is closely related to the natural ligand of the polypeptide, e.g., a fragment of the ligand, or a natural substrate, a ligand, a structural or functional mimetic. (See, Coligan et al., Current Protocols in Immunology 1(2)/Chapter 5 (1991).) Similarly, the molecule can be closely related to the natural receptor to which the polypeptide binds, or at least, a fragment of the receptor capable of being bound by the polypeptide (e.g., active site). In either case, the molecule can be rationally designed using known techniques.

Preferably, the screening for these molecules involves producing appropriate cells which express the polypeptide, either as a secreted protein or on the cell membrane. Preferred cells include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing the polypeptide (or cell membrane containing the expressed polypeptide) are then preferably contacted with a test compound potentially containing the molecule to observe binding, stimulation, or inhibition of activity of either the polypeptide or the molecule.

The assay may simply test binding of a candidate compound to the polypeptide, wherein binding is detected by a label, or in an assay involving competition with a labeled competitor. Further, the assay may test whether the candidate compound results in a signal generated by binding to the polypeptide.

Alternatively, the assay can be carried out using cell-free preparations, polypeptide/molecule affixed to a solid support, chemical libraries, or natural product mixtures. The assay may also simply comprise the steps of mixing a candidate compound with a solution containing a polypeptide, measuring polypeptide/molecule activity or binding, and comparing the polypeptide/molecule activity or binding to a standard.

Preferably, an ELISA assay can measure polypeptide level or activity in a sample (e.g., biological sample) using a monoclonal or polyclonal antibody. The

antibody can measure polypeptide level or activity by either binding, directly or indirectly, to the polypeptide or by competing with the polypeptide for a substrate.

All of these above assays can be used as diagnostic or prognostic markers. The molecules discovered using these assays can be used to treat disease or to bring about a particular result in a patient (e.g., blood vessel growth) by activating or inhibiting the polypeptide/molecule. Moreover, the assays can discover agents which may inhibit or enhance the production of the polypeptide from suitably manipulated cells or tissues.

Therefore, the invention includes a method of identifying compounds which bind to a polypeptide of the invention comprising the steps of: (a) incubating a candidate binding compound with a polypeptide of the invention; and (b) determining if binding has occurred. Moreover, the invention includes a method of identifying agonists/antagonists comprising the steps of: (a) incubating a candidate compound with a polypeptide of the invention, (b) assaying a biological activity, and (b) determining if a biological activity of the polypeptide has been altered.

15

#### Other Activities

A polypeptide or polynucleotide of the present invention may also increase or decrease the differentiation or proliferation of embryonic stem cells, besides, as discussed above, hematopoietic lineage.

A polypeptide or polynucleotide of the present invention may also be used to modulate mammalian characteristics, such as body height, weight, hair color, eye color, skin, percentage of adipose tissue, pigmentation, size, and shape (e.g., cosmetic surgery). Similarly, a polypeptide or polynucleotide of the present invention may be used to modulate mammalian metabolism affecting catabolism, anabolism, processing, utilization, and storage of energy.

25

A polypeptide or polynucleotide of the present invention may be used to change a mammal's mental state or physical state by influencing biorhythms, circadian rhythms, depression (including depressive disorders), tendency for violence, tolerance for pain, reproductive capabilities (preferably by Activin or Inhibin-like activity), hormonal or endocrine levels, appetite, libido, memory, stress, or other cognitive qualities.

30

A polypeptide or polynucleotide of the present invention may also be used as a food additive or preservative, such as to increase or decrease storage capabilities, fat content, lipid, protein, carbohydrate, vitamins, minerals, cofactors or other nutritional components.

35

#### Other Preferred Embodiments

Other preferred embodiments of the claimed invention include an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 50 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1.

5

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Clone Sequence and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

10

Also preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the Start Codon and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

15

Similarly preferred is a nucleic acid molecule wherein said sequence of contiguous nucleotides is included in the nucleotide sequence of SEQ ID NO:X in the range of positions beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

20

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 150 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

25

Further preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least about 500 contiguous nucleotides in the nucleotide sequence of SEQ ID NO:X.

Table 1.

A further preferred embodiment is a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the nucleotide sequence of SEQ ID NO:X beginning with the nucleotide at about the position of the 5' Nucleotide of the First Amino Acid of the Signal Peptide and ending with the nucleotide at about the position of the 3' Nucleotide of the Clone Sequence as defined for SEQ ID NO:X in Table 1.

30

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence of SEQ ID NO:X.

35



Also preferred is an isolated nucleic acid molecule which hybridizes under stringent hybridization conditions to a nucleic acid molecule, wherein said nucleic acid molecule which hybridizes does not hybridize under stringent hybridization conditions to a nucleic acid molecule having a nucleotide sequence consisting of only A residues or of only T residues.

Also preferred is a composition of matter comprising a DNA molecule which comprises a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the material deposited with the American Type Culture Collection and given the ATCC Deposit Number shown in Table 1 for said cDNA Clone Identifier.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in the nucleotide sequence of a human cDNA clone identified by a cDNA Clone Identifier in Table 1, which DNA molecule is contained in the deposit given the ATCC Deposit Number shown in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said sequence of at least 50 contiguous nucleotides is included in the nucleotide sequence of the complete open reading frame sequence encoded by said human cDNA clone.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 150 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to sequence of at least 500 contiguous nucleotides in the nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to the complete nucleotide sequence encoded by said human cDNA clone.

A further preferred embodiment is a method for detecting in a biological sample a nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing a nucleotide sequence of at least one nucleic acid molecule in said sample with a sequence selected from said group and determining

whether the sequence of said nucleic acid molecule in said sample is at least 95% identical to said selected sequence.

Also preferred is the above method wherein said step of comparing sequences comprises determining the extent of nucleic acid hybridization between nucleic acid molecules in said sample and a nucleic acid molecule comprising said sequence selected from said group. Similarly, also preferred is the above method wherein said step of comparing sequences is performed by comparing the nucleotide sequence determined from a nucleic acid molecule in said sample with said sequence selected from said group. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

A further preferred embodiment is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting nucleic acid molecules in said sample, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for identifying the species, tissue or cell type of a biological sample can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject nucleic acid molecules, if any, comprising a nucleotide sequence that is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

The method for diagnosing a pathological condition can comprise a step of detecting nucleic acid molecules comprising a nucleotide sequence in a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from said group.

Also preferred is a composition of matter comprising isolated nucleic acid molecules wherein the nucleotide sequences of said nucleic acid molecules comprise a panel of at least two nucleotide sequences, wherein at least one sequence in said panel is at least 95% identical to a sequence of at least 50 contiguous nucleotides in a sequence selected from the group consisting of: a nucleotide sequence of SEQ ID NO:X wherein X is any integer as defined in Table 1; and a nucleotide sequence encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The nucleic acid molecules can comprise DNA molecules or RNA molecules.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1.

Also preferred is a polypeptide, wherein said sequence of contiguous amino acids is included in the amino acid sequence of SEQ ID NO:Y in the range of positions beginning with the residue at about the position of the First Amino Acid of the Secreted Portion and ending with the residue at about the Last Amino Acid of the Open Reading Frame as set forth for SEQ ID NO:Y in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the complete amino acid sequence of SEQ ID NO:Y.

Further preferred is an isolated polypeptide comprising an amino acid sequence at least 90% identical to a sequence of at least about 10 contiguous amino acids in the complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is a polypeptide wherein said sequence of contiguous amino acids is included in the amino acid sequence of a secreted portion of the secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 30 contiguous amino acids in the

amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence of at least about 100 contiguous amino acids in the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated polypeptide comprising an amino acid sequence at least 95% identical to the amino acid sequence of the secreted portion of the protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is an isolated antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method for detecting in a biological sample a polypeptide comprising an amino acid sequence which is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1; which method comprises a step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group and determining whether the sequence of said polypeptide molecule in said sample is at least 90% identical to said sequence of at least 10 contiguous amino acids.

Also preferred is the above method wherein said step of comparing an amino acid sequence of at least one polypeptide molecule in said sample with a sequence selected from said group comprises determining the extent of specific binding of polypeptides in said sample to an antibody which binds specifically to a polypeptide comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an

amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method wherein said step of comparing sequences is performed by comparing the amino acid sequence determined from a polypeptide molecule in said sample with said sequence selected from said group.

Also preferred is a method for identifying the species, tissue or cell type of a biological sample which method comprises a step of detecting polypeptide molecules in said sample, if any, comprising an amino acid sequence that is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is the above method for identifying the species, tissue or cell type of a biological sample, which method comprises a step of detecting polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the above group.

Also preferred is a method for diagnosing in a subject a pathological condition associated with abnormal structure or expression of a gene encoding a secreted protein identified in Table 1, which method comprises a step of detecting in a biological sample obtained from said subject polypeptide molecules comprising an amino acid sequence in a panel of at least two amino acid sequences, wherein at least one sequence in said panel is at least 90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

In any of these methods, the step of detecting said polypeptide molecules includes using an antibody.

Also preferred is an isolated nucleic acid molecule comprising a nucleotide sequence which is at least 95% identical to a nucleotide sequence encoding a polypeptide wherein said polypeptide comprises an amino acid sequence that is at least

90% identical to a sequence of at least 10 contiguous amino acids in a sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Also preferred is an isolated nucleic acid molecule, wherein said nucleotide sequence encoding a polypeptide has been optimized for expression of said polypeptide in a prokaryotic host.

Also preferred is an isolated nucleic acid molecule, wherein said polypeptide comprises an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y wherein Y is any integer as defined in Table 1; and a complete amino acid sequence of a secreted protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1.

Further preferred is a method of making a recombinant vector comprising inserting any of the above isolated nucleic acid molecule into a vector. Also preferred is the recombinant vector produced by this method. Also preferred is a method of making a recombinant host cell comprising introducing the vector into a host cell, as well as the recombinant host cell produced by this method.

Also preferred is a method of making an isolated polypeptide comprising culturing this recombinant host cell under conditions such that said polypeptide is expressed and recovering said polypeptide. Also preferred is this method of making an isolated polypeptide, wherein said recombinant host cell is a eukaryotic cell and said polypeptide is a secreted portion of a human secreted protein comprising an amino acid sequence selected from the group consisting of: an amino acid sequence of SEQ ID NO:Y beginning with the residue at the position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y wherein Y is an integer set forth in Table 1 and said position of the First Amino Acid of the Secreted Portion of SEQ ID NO:Y is defined in Table 1 and an amino acid sequence of a secreted portion of a protein encoded by a human cDNA clone identified by a cDNA Clone Identifier in Table 1 and contained in the deposit with the ATCC Deposit Number shown for said cDNA clone in Table 1. The isolated polypeptide produced by this method is also preferred.

Also preferred is a method of treatment of an individual in need of an increased level of a secreted protein activity, which method comprises administering to such an individual a pharmaceutical composition comprising an amount of an isolated

polypeptide, polynucleotide, or antibody of the claimed invention effective to increase the level of said protein activity in said individual.

Having generally described the invention, the same will be more readily understood by reference to the following examples, which are provided by way of illustration and are not intended as limiting.

### Examples

#### Example 1: Isolation of a Selected cDNA Clone From the Deposited Sample

- 10 Each cDNA clone in a cited ATCC deposit is contained in a plasmid vector. Table 1 identifies the vectors used to construct the cDNA library from which each clone was isolated. In many cases, the vector used to construct the library is a phage vector from which a plasmid has been excised. The table immediately below correlates the related plasmid for each phage vector used in constructing the cDNA library. For example, where a particular clone is identified in Table 1 as being isolated in the vector "Lambda Zap," the corresponding deposited clone is in "pBluescript."

Vector Used to Construct Library	Corresponding Deposited Plasmid
Lambda Zap	pBluescript (pBS)
Uni-Zap XR	pBluescript (pBS)
Zap Express	pBK
lambid BA	plasmid BA
pSport1	pSport1
pCMVSPORT 2.0	pCMVSPORT 2.0
pCMVSPORT 3.0	pCMVSPORT 3.0
pCR <sup>2.1</sup>	pCR <sup>2.1</sup>

- 25 Vectors Lambda Zap (U.S. Patent Nos. 5,128,256 and 5,286,636), Uni-Zap XR (U.S. Patent Nos. 5,128,256 and 5,286,636), Zap Express (U.S. Patent Nos. 5,128,256 and 5,286,636), pBluescript (pBS) (Short, J. M. et al., Nucleic Acids Res. 16:7583-7600 (1988); Alting-Mees, M. A. and Short, J. M., Nucleic Acids Res. 17:9494 (1989)) and pBK (Alting-Mees, M. A. et al., Strategies 5:58-61 (1992)) are commercially available from Stratagene Cloning Systems, Inc., 11011 N. Torrey Pines Road, La Jolla, CA, 92037. pBS contains an ampicillin resistance gene and pBK contains a neomycin resistance gene. Both can be transformed into E. coli strain XL-1 Blue, also available from Stratagene. pBS comes in 4 forms SK+, SK-, KS+ and KS-. The S and K refers to the orientation of the polylinker to the T7 and T3 primer sequences which flank the polylinker region ("S" is for SacI and "K" is for KpnI which are the first sites on each respective end of the linker). "+" or "-" refer to the orientation

of the f1 origin of replication ("ori"), such that in one orientation, single stranded rescue initiated from the f1 ori generates sense strand DNA and in the other, antisense.

- 5 Vectors pSport1, pCMVSPORT 2.0 and pCMVSPORT 3.0, were obtained from Life Technologies, Inc., P. O. Box 6009, Gaithersburg, MD 20897. All Sport vectors contain an ampicillin resistance gene and may be transformed into E. coli strain DH10B, also available from Life Technologies. (See, for instance, Gruber, C. E., et al., Focus 15:59 (1993).) Vector lalfmid BA (Benito Soares, Columbia University, NY) contains an ampicillin resistance gene and can be transformed into E. coli strain XL-1 Blue. Vector pCR<sup>2.1</sup>, which is available from Invitrogen, 1600 Faraday Avenue, Carlsbad, CA 92008, contains an ampicillin resistance gene and may be transformed into E. coli strain DH10B, available from Life Technologies. (See, for instance, Clontech, J. M., Nuc. Acids Res. 16:9677-9686 (1988) and Mead, D. et al., Bio/Technology 9: (1991).) Preferably, a polynucleotide of the present invention does not comprise the phage vector sequences identified for the particular clone in Table 1, as well as the corresponding plasmid vector sequences designated above.

- 15 The deposited material in the sample assigned the ATCC Deposit Number cited in Table 1 for any given cDNA clone also may contain one or more additional plasmids, each comprising a cDNA clone different from that given clone. Thus, deposits sharing the same ATCC Deposit Number contain at least a plasmid for each cDNA clone identified in Table 1. Typically, each ATCC deposit sample cited in Table 1 comprises a mixture of approximately equal amounts (by weight) of about 50 plasmid DNAs, each containing a different cDNA clone; but such a deposit sample may include plasmids for more or less than 50 cDNA clones, up to about 500 cDNA clones.

- 25 Two approaches can be used to isolate a particular clone from the deposited sample of plasmid DNAs cited for that clone in Table 1. First, a plasmid is directly isolated by screening the clones using a polynucleotide probe corresponding to SEQ ID NO:X.

Particularly, a specific polynucleotide with 30-40 nucleotides is synthesized using an Applied Biosystems DNA synthesizer according to the sequence reported.

- 30 The oligonucleotide is labeled, for instance, with <sup>32</sup>P-γ-ATP using T4 polynucleotide kinase and purified according to routine methods. (E.g., Maniatis et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Press, Cold Spring, NY (1982).) The plasmid mixture is transformed into a suitable host, as indicated above (such as XL-1 Blue (Stratagene)) using techniques known to those of skill in the art, such as those provided by the vector supplier or in related publications or patents cited above. The transformants are plated on 1.5% agar plates (containing the appropriate selection

agent, e.g., ampicillin) to a density of about 150 transformants (colonies) per plate. These plates are screened using Nylon membranes according to routine methods for bacterial colony screening (e.g., Sambrook et al., Molecular Cloning: A Laboratory Manual, 2nd Edit., (1989), Cold Spring Harbor Laboratory Press, pages 1.93 to 1.104), or other techniques known to those of skill in the art.

Alternatively, two primers of 17-20 nucleotides derived from both ends of the SEQ ID NO:X (i.e., within the region of SEQ ID NO:X bounded by the 5' NT and the 3' NT of the clone defined in Table 1) are synthesized and used to amplify the desired cDNA using the deposited cDNA plasmid as a template. The polymerase chain reaction is carried out under routine conditions, for instance, in 25  $\mu$ l of reaction mixture with 0.5  $\mu$ g of the above cDNA template. A convenient reaction mixture is 1.5-5 mM  $MgCl_2$ , 0.01% (w/v) gelatin, 20  $\mu$ M each of dATP, dCTP, dGTP, dTTP, 25 pmol of each primer and 0.25 Unit of Taq polymerase. Thirty five cycles of PCR (denaturation at 94°C for 1 min; annealing at 55°C for 1 min; elongation at 72°C for 1 min) are performed with a Perkin-Elmer Cetus automated thermal cycler. The amplified product is analyzed by agarose gel electrophoresis and the DNA band with expected molecular weight is excised and purified. The PCR product is verified to be the selected sequence by subcloning and sequencing the DNA product.

Several methods are available for the identification of the 5' or 3' non-coding portions of a gene which may not be present in the deposited clone. These methods include but are not limited to, filter probing, clone enrichment using specific probes, and protocols similar or identical to 5' and 3' "RACE" protocols which are well known in the art. For instance, a method similar to 5' RACE is available for generating the missing 5' end of a desired full-length transcript. (Fromont-Racine et al., Nucleic Acids Res. 21(7):1683-1684 (1993).)

Briefly, a specific RNA oligonucleotide is ligated to the 5' ends of a population of RNA presumably containing full-length gene RNA transcripts. A primer set containing a primer specific to the ligated RNA oligonucleotide and a primer specific to a known sequence of the gene of interest is used to PCR amplify the 5' portion of the desired full-length gene. This amplified product may then be sequenced and used to generate the full length gene.

This above method starts with total RNA isolated from the desired source, although poly-A+ RNA can be used. The RNA preparation can then be treated with phosphatase if necessary to eliminate 5' phosphate groups on degraded or damaged RNA which may interfere with the later RNA ligase step. The phosphatase should then be inactivated and the RNA treated with tobacco acid pyrophosphatase in order to

remove the cap structure present at the 5' ends of messenger RNAs. This reaction leaves a 5' phosphate group at the 5' end of the cap cleaved RNA which can then be ligated to an RNA oligonucleotide using T4 RNA ligase.

This modified RNA preparation is used as a template for first strand cDNA synthesis using a gene specific oligonucleotide. The first strand synthesis reaction is used as a template for PCR amplification of the desired 5' end using a primer specific to the ligated RNA oligonucleotide and a primer specific to the known sequence of the gene of interest. The resultant product is then sequenced and analyzed to confirm that the 5' end sequence belongs to the desired gene.

#### **Example 2: Isolation of Genomic Clones Corresponding to a Polynucleotide**

A human genomic PI library (Genomic Systems, Inc.) is screened by PCR using primers selected for the cDNA sequence corresponding to SEQ ID NO:X., according to the method described in Example 1. (See also, Sambrook.)

#### **Example 3: Tissue Distribution of Polypeptide**

Tissue distribution of mRNA expression of polynucleotides of the present invention is determined using protocols for Northern blot analysis, described by, among others, Sambrook et al. For example, a cDNA probe produced by the method described in Example 1 is labeled with  $P^{32}$  using the rediprimer™ DNA labeling system (Amersham Life Science), according to manufacturer's instructions. After labeling, the probe is purified using CHROMA SPIN-100™ column (Clontech Laboratories, Inc.), according to manufacturer's protocol number PT1200-1. The purified labeled probe is then used to examine various human tissues for mRNA expression.

Multiple Tissue Northern (MTN) blots containing various human tissues (H) or human immune system tissues (IM) (Clontech) are examined with the labeled probe using ExpressHyb™ hybridization solution (Clontech) according to manufacturer's protocol number PT1190-1. Following hybridization and washing, the blots are mounted and exposed to film at -70°C overnight, and the films developed according to standard procedures.

#### **Example 4: Chromosomal Mapping of the Polynucleotides**

An oligonucleotide primer set is designed according to the sequence at the 5' end of SEQ ID NO:X. This primer preferably spans about 100 nucleotides. This primer set is then used in a polymerase chain reaction under the following set of

conditions : 30 seconds, 95°C; 1 minute, 56°C; 1 minute, 70°C. This cycle is repeated 32 times followed by one 5 minute cycle at 70°C. Human, mouse, and hamster DNA is used as template in addition to a somatic cell hybrid panel containing individual chromosomes or chromosome fragments (Bios, Inc). The reactions is analyzed on either 8% polyacrylamide gels or 3.5 % agarose gels. Chromosome mapping is determined by the presence of an approximately 100 bp PCR fragment in the particular somatic cell hybrid.

#### Example 5: Bacterial Expression of a Polypeptide

A polynucleotide encoding a polypeptide of the present invention is amplified using PCR oligonucleotide primers corresponding to the 5' and 3' ends of the DNA sequence, as outlined in Example 1, to synthesize insertion fragments. The primers used to amplify the cDNA insert should preferably contain restriction sites, such as BamHI and XbaI, at the 5' end of the primers in order to clone the amplified product into the expression vector. For example, BamHI and XbaI correspond to the restriction enzyme sites on the bacterial expression vector pQE-9. (Qiagen, Inc., Chatsworth, CA). This plasmid vector encodes antibiotic resistance (Amp<sup>r</sup>), a bacterial origin of replication (ori), an IPTG-regulatable promoter/operator (P/O), a ribosome binding site (RBS), a 6-histidine tag (6-His), and restriction enzyme cloning sites.

The pQE-9 vector is digested with BamHI and XbaI and the amplified fragment is ligated into the pQE-9 vector maintaining the reading frame initiated at the bacterial RBS. The ligation mixture is then used to transform the E. coli strain M15/rep4 (Qiagen, Inc.) which contains multiple copies of the plasmid pREP4, which expresses the lacI repressor and also confers kanamycin resistance (Kan<sup>r</sup>). Transformants are identified by their ability to grow on LB plates and ampicillin/kanamycin resistant colonies are selected. Plasmid DNA is isolated and confirmed by restriction analysis.

Clones containing the desired constructs are grown overnight (O/N) in liquid culture in LB media supplemented with both Amp (100 ug/ml) and Kan (25 ug/ml). The ON culture is used to inoculate a large culture at a ratio of 1:100 to 1:250. The cells are grown to an optical density 600 (O.D.<sub>600</sub>) of between 0.4 and 0.6. IPTG (Isopropyl-B-D-thiogalactopyranoside) is then added to a final concentration of 1 mM. IPTG induces by inactivating the lacI repressor, clearing the P/O leading to increased gene expression.

Cells are grown for an extra 3 to 4 hours. Cells are then harvested by centrifugation (20 mins at 6000Xg). The cell pellet is solubilized in the chaotropic

agent 6 Molar Guanidine HCl by stirring for 3-4 hours at 4°C. The cell debris is removed by centrifugation, and the supernatant containing the polypeptide is loaded onto a nickel-nitrilo-tri-acetic acid ("Ni-NTA") affinity resin column (available from QIAGEN, Inc., *supra*). Proteins with a 6 x His tag bind to the Ni-NTA resin with high affinity and can be purified in a simple one-step procedure (for details see: The QIAexpressionist (1995) QIAGEN, Inc., *supra*).

Briefly, the supernatant is loaded onto the column in 6 M guanidine-HCl, pH 8, the column is first washed with 10 volumes of 6 M guanidine-HCl, pH 8, then washed with 10 volumes of 6 M guanidine-HCl pH 6, and finally the polypeptide is eluted with 6 M guanidine-HCl, pH 5.

The purified protein is then renatured by dialyzing it against phosphate-buffered saline (PBS) or 50 mM Na-acetate, pH 6 buffer plus 200 mM NaCl. Alternatively, the protein can be successfully refolded while immobilized on the Ni-NTA column. The recommended conditions are as follows: renature using a linear 6M-1M urea gradient in 500 mM NaCl, 20% glycerol, 20 mM Tris/HCl pH 7.4, containing protease inhibitors. The renaturation should be performed over a period of 1.5 hours or more. After renaturation the proteins are eluted by the addition of 250 mM imidazole. Imidazole is removed by a final dialyzing step against PBS or 50 mM sodium acetate pH 6 buffer plus 200 mM NaCl. The purified protein is stored at 4°C or frozen at -80°C.

In addition to the above expression vector, the present invention further includes an expression vector comprising phage operator and promoter elements operatively linked to a polynucleotide of the present invention, called pHE4a. (ATCC Accession Number XXXXXX.) This vector contains: 1) a neomycinphosphotransferase gene as a selection marker, 2) an E. coli origin of replication, 3) a TS phage promoter sequence, 4) two lac operator sequences, 5) a Shine-Delgarno sequence, and 6) the lactose operon repressor gene (lacIq). The origin of replication (oriC) is derived from pUC19 (LTI, Gaithersburg, MD). The promoter sequence and operator sequences are made synthetically.

DNA can be inserted into the pHEa by restricting the vector with NdeI and XbaI, BamHI, XhoI, or Asp718, running the restricted product on a gel, and isolating the larger fragment (the stuffer fragment should be about 310 base pairs). The DNA insert is generated according to the PCR protocol described in Example 1, using PCR primers having restriction sites for NdeI (5' primer) and XbaI, BamHI, XhoI, or Asp718 (3' primer). The PCR insert is gel purified and restricted with compatible enzymes. The insert and vector are ligated according to standard protocols.

The engineered vector could easily be substituted in the above protocol to express protein in a bacterial system.

#### **Example 6: Purification of a Polypeptide from an Inclusion Body**

The following alternative method can be used to purify a polypeptide expressed in *E. coli* when it is present in the form of inclusion bodies. Unless otherwise specified, all of the following steps are conducted at 4-10°C.

Upon completion of the production phase of the *E. coli* fermentation, the cell culture is cooled to 4-10°C and the cells harvested by continuous centrifugation at 15,000 rpm (Heraeus Sepatech). On the basis of the expected yield of protein per unit weight of cell paste and the amount of purified protein required, an appropriate amount of cell paste, by weight, is suspended in a buffer solution containing 100 mM Tris, 50 mM EDTA, pH 7.4. The cells are dispersed to a homogeneous suspension using a high shear mixer.

The cells are then lysed by passing the solution through a microfluidizer (Microfluidics, Corp. or APV Gaulin, Inc.) twice at 4000-6000 psi. The homogenate is then mixed with NaCl solution to a final concentration of 0.5 M NaCl, followed by centrifugation at 7000 xg for 15 min. The resultant pellet is washed again using 0.5M NaCl, 100 mM Tris, 50 mM EDTA, pH 7.4.

The resulting washed inclusion bodies are solubilized with 1.5 M guanidine hydrochloride (GuHCl) for 2-4 hours. After 7000 xg centrifugation for 15 min., the pellet is discarded and the polypeptide containing supernatant is incubated at 4°C overnight to allow further GuHCl extraction.

Following high speed centrifugation (30,000 xg) to remove insoluble particles, the GuHCl solubilized protein is refolded by quickly mixing the GuHCl extract with 20 volumes of buffer containing 50 mM sodium, pH 4.5, 150 mM NaCl, 2 mM EDTA by vigorous stirring. The refolded diluted protein solution is kept at 4°C without mixing for 12 hours prior to further purification steps.

To clarify the refolded polypeptide solution, a previously prepared tangential filtration unit equipped with 0.16 µm membrane filter with appropriate surface area (e.g., Filtron), equilibrated with 40 mM sodium acetate, pH 6.0 is employed. The filtered sample is loaded onto a cation exchange resin (e.g., Poros HS-50, Perseptive Biosystems). The column is washed with 40 mM sodium acetate, pH 6.0 and eluted with 250 mM, 500 mM, 1000 mM, and 1500 mM NaCl in the same buffer, in a

stepwise manner. The absorbance at 280 nm of the effluent is continuously monitored. Fractions are collected and further analyzed by SDS-PAGE.

Fractions containing the polypeptide are then pooled and mixed with 4 volumes of water. The diluted sample is then loaded onto a previously prepared set of tandem columns of strong anion (Poros HQ-50, Perseptive Biosystems) and weak anion (Poros CM-20, Perseptive Biosystems) exchange resins. The columns are equilibrated with 40 mM sodium acetate, pH 6.0. Both columns are washed with 40 mM sodium acetate, pH 6.0, 200 mM NaCl. The CM-20 column is then eluted using a 10 column volume linear gradient ranging from 0.2 M NaCl, 50 mM sodium acetate, pH 6.0 to 1.0 M NaCl, 50 mM sodium acetate, pH 6.5. Fractions are collected under constant  $A_{280}$  monitoring of the effluent. Fractions containing the polypeptide (determined, for instance, by 16% SDS-PAGE) are then pooled.

The resultant polypeptide should exhibit greater than 95% purity after the above refolding and purification steps. No major contaminant bands should be observed from Commassie blue stained 16% SDS-PAGE gel when 5 µg of purified protein is loaded. The purified protein can also be tested for endotoxin/LPS contamination, and typically the LPS content is less than 0.1 ng/ml according to LAL assays.

#### **Example 7: Cloning and Expression of a Polypeptide in a Baculovirus Expression System**

In this example, the plasmid shuttle vector pA2 is used to insert a polynucleotide into a baculovirus to express a polypeptide. This expression vector contains the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) followed by convenient restriction sites such as BamHI, Xba I and Asp718. The polyadenylation site of the simian virus 40 ("SV40") is used for efficient polyadenylation. For easy selection of recombinant virus, the plasmid contains the beta-galactosidase gene from *E. coli* under control of a weak Drosophila promoter in the same orientation, followed by the polyadenylation signal of the polyhedrin gene. The inserted genes are flanked on both sides by viral sequences for cell-mediated homologous recombination with wild-type viral DNA to generate a viable virus that express the cloned polynucleotide.

Many other baculovirus vectors can be used in place of the vector above, such as pAc373, pVL941, and pAcIM1, as one skilled in the art would readily appreciate, as long as the construct provides appropriately located signals for transcription, translation, secretion and the like, including a signal peptide and an in-frame AUG as

required. Such vectors are described, for instance, in Luckow et al., *Virology* 170:31-39 (1989).

Specifically, the cDNA sequence contained in the deposited clone, including the AUG initiation codon and the naturally associated leader sequence identified in Table 1, is amplified using the PCR protocol described in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the pA2 vector does not need a second signal peptide. Alternatively, the vector can be modified (pA2 GP) to include a baculovirus leader sequence, using the standard methods described in Summers et al., "A Manual of Methods for Baculovirus Vectors and Insect Cell Culture Procedures," Texas Agricultural Experimental Station Bulletin No. 1555 (1987).

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

The plasmid is digested with the corresponding restriction enzymes and optionally, can be dephosphorylated using calf intestinal phosphatase, using routine procedures known in the art. The DNA is then isolated from a 1% agarose gel using a commercially available kit ("GeneClean" BIO 101 Inc., La Jolla, Ca.).

The fragment and the dephosphorylated plasmid are ligated together with T4 DNA ligase. *E. coli* HB101 or other suitable *E. coli* hosts such as XL-1 Blue (Stratagene Cloning Systems, La Jolla, CA) cells are transformed with the ligation mixture and spread on culture plates. Bacteria containing the plasmid are identified by digesting DNA from individual colonies and analyzing the digestion product by gel electrophoresis. The sequence of the cloned fragment is confirmed by DNA sequencing.

Five µg of a plasmid containing the polynucleotide is co-transfected with 1.0 µg of a commercially available linearized baculovirus DNA ("BaculoGold™ baculovirus DNA", Pharmingen, San Diego, CA), using the lipofection method described by Felgner et al., *Proc. Natl. Acad. Sci. USA* 84:7413-7417 (1987). One µg of BaculoGold™ virus DNA and 5 µg of the plasmid are mixed in a sterile well of a microtiter plate containing 50 µl of serum-free Grace's medium (Life Technologies Inc., Gaithersburg, MD). Afterwards, 10 µl Lipofectin plus 90 µl Grace's medium are added, mixed and incubated for 15 minutes at room temperature. Then the transfection mixture is added drop-wise to Sf9 insect cells (ATCC CRL 1711) seeded in a 35 mm tissue culture plate with 1 ml Grace's medium without serum. The plate is then incubated for 5 hours at 27° C. The transfection solution is then removed from the plate and 1 ml of Grace's insect medium supplemented with 10% fetal calf serum is added. Cultivation is then continued at 27° C for four days.

After four days the supernatant is collected and a plaque assay is performed, as described by Summers and Smith, *supra*. An agarose gel with "Blue Gel" (Life Technologies Inc., Gaithersburg) is used to allow easy identification and isolation of gal-expressing clones, which produce blue-stained plaques. (A detailed description of a "plaque assay" of this type can also be found in the user's guide for insect cell culture and baculovirology distributed by Life Technologies Inc., Gaithersburg, page 9-10.) After appropriate incubation, blue stained plaques are picked with the tip of a micropipettor (e.g., Eppendorf). The agar containing the recombinant viruses is then resuspended in a microcentrifuge tube containing 200 µl of Grace's medium and the suspension containing the recombinant baculovirus is used to infect Sf9 cells seeded in 35 mm dishes. Four days later the supernatants of these culture dishes are harvested and then they are stored at 4° C.

To verify the expression of the polypeptide, Sf9 cells are grown in Grace's medium supplemented with 10% heat-inactivated FBS. The cells are infected with the recombinant baculovirus containing the polynucleotide at a multiplicity of infection ("MOI") of about 2. If radiolabeled proteins are desired, 6 hours later the medium is removed and is replaced with SF900 II medium minus methionine and cysteine (available from Life Technologies Inc., Rockville, MD). After 42 hours, 5 µCi of <sup>35</sup>S-methionine and 5 µCi <sup>35</sup>S-cysteine (available from Amersham) are added. The cells are further incubated for 16 hours and then are harvested by centrifugation. The proteins in the supernatant as well as the intracellular proteins are analyzed by SDS-PAGE followed by autoradiography (if radiolabeled).

Microsequencing of the amino acid sequence of the amino terminus of purified protein may be used to determine the amino terminal sequence of the produced protein.

## Example 8: Expression of a Polypeptide in Mammalian Cells

The polypeptide of the present invention can be expressed in a mammalian cell. A typical mammalian expression vector contains a promoter element, which mediates the initiation of transcription of mRNA, a protein coding sequence, and signals required for the termination of transcription and polyadenylation of the transcript. Additional elements include enhancers, Kozak sequences and intervening sequences flanked by donor and acceptor sites for RNA splicing. Highly efficient transcription is achieved with the early and late promoters from SV40, the long terminal repeats (LTRs) from Retroviruses, e.g., RSV, HTLV, HIV1 and the early promoter of the cytomegalovirus (CMV). However, cellular elements can also be used (e.g., the human actin promoter).

Suitable expression vectors for use in practicing the present invention include, for example, vectors such as pSVL and pMSG (Pharmacia, Uppsala, Sweden).



5 pRSVcat (ATCC 37152), pSV2dhfr (ATCC 37146), pBC12MI (ATCC 67109), pCMVSPORT 2.0, and pCMVSPORT 3.0. Mammalian host cells that could be used include, human Hela, 293, H9 and Jurkat cells, mouse NIH3T3 and C127 cells, Cos 1, Cos 7 and CV1, quail QG1-3 cells, mouse L cells and Chinese hamster ovary (CHO) cells.

Alternatively, the polypeptide can be expressed in stable cell lines containing the polynucleotide integrated into a chromosome. The co-transfection with a selectable marker such as dhfr, gpt, neomycin, hygromycin allows the identification and isolation of the transfected cells.

10 The transfected gene can also be amplified to express large amounts of the encoded protein. The DHFR (dihydrofolate reductase) marker is useful in developing cell lines that carry several hundred or even several thousand copies of the gene of interest. (See, e.g., Alt, F. W., et al., J. Biol. Chem. 253:1357-1370 (1978); Hamlin, J. L. and Ma, C., Biochem. et Biophys. Acta, 1097:107-143 (1990); Page, M. J. and Sydenham, M. A., Biotechnology 9:64-68 (1991).) Another useful selection marker is the enzyme glutamine synthase (GS) (Murphy et al., Biochem J. 227:277-279 (1991); Bebbington et al., Bio/Technology 10:169-175 (1992).) Using these markers, the mammalian cells are grown in selective medium and the cells with the highest resistance are selected. These cell lines contain the amplified gene(s) integrated into a chromosome. Chinese hamster ovary (CHO) and NSO cells are often used for the production of proteins.

Derivatives of the plasmid pSV2-dhfr (ATCC Accession No. 37146), the expression vectors pC4 (ATCC Accession No. 209646) and pC6 (ATCC Accession No. 209647) contain the strong promoter (LTR) of the Rous Sarcoma Virus (Cullen et al., Molecular and Cellular Biology, 438-447 (March, 1985)) plus a fragment of the CMV-enhancer (Boshart et al., Cell 41:521-530 (1985)). Multiple cloning sites, e.g., with the restriction enzyme cleavage sites BamHI, XbaI and Asp718, facilitate the cloning of the gene of interest. The vectors also contain the 3' intron, the polyadenylation and termination signal of the rat preproinsulin gene, and the mouse DHFR gene under control of the SV40 early promoter.

Specifically, the plasmid pC6, for example, is digested with appropriate restriction enzymes and then dephosphorylated using calf intestinal phosphates by procedures known in the art. The vector is then isolated from a 1% agarose gel.

35 A polynucleotide of the present invention is amplified according to the protocol outlined in Example 1. If the naturally occurring signal sequence is used to produce the secreted protein, the vector does not need a second signal peptide. Alternatively, if the

naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

The amplified fragment is isolated from a 1% agarose gel using a commercially available kit ("GeneClean," BIO 101 Inc., La Jolla, Ca.). The fragment then is digested with appropriate restriction enzymes and again purified on a 1% agarose gel.

5 The amplified fragment is then digested with the same restriction enzyme and purified on a 1% agarose gel. The isolated fragment and the dephosphorylated vector are then ligated with T4 DNA ligase. *E. coli* HB101 or XL-1 Blue cells are then transformed and bacteria are identified that contain the fragment inserted into plasmid pC6 using, for instance, restriction enzyme analysis.

10 Chinese hamster ovary cells lacking an active DHFR gene is used for transfection. Five  $\mu$ g of the expression plasmid pC6 is cotransfected with 0.5  $\mu$ g of the plasmid pSVneo using lipofectin (Felgner et al., *supra*). The plasmid pSV2-neo contains a dominant selectable marker, the *neo* gene from Tn5 encoding an enzyme that confers resistance to a group of antibiotics including G418. The cells are seeded in alpha minus MEM supplemented with 1 mg/ml G418. After 2 days, the cells are trypsinized and seeded in hybridoma cloning plates (Greiner, Germany) in alpha minus MEM supplemented with 10, 25, or 50 ng/ml of methotrexate plus 1 mg/ml G418. After about 10-14 days single clones are trypsinized and then seeded in 6-well petri dishes or 10 ml flasks using different concentrations of methotrexate (50 nM, 100 nM, 200 nM, 400 nM, 800 nM). Clones growing at the highest concentrations of methotrexate are then transferred to new 6-well plates containing even higher concentrations of methotrexate (1  $\mu$ M, 2  $\mu$ M, 5  $\mu$ M, 10 mM, 20 mM). The same procedure is repeated until clones are obtained which grow at a concentration of 100 - 200  $\mu$ M. Expression of the desired gene product is analyzed, for instance, by SDS-PAGE and Western blot or by reversed phase HPLC analysis.

### Example 9: Protein Fusions

The polypeptides of the present invention are preferably fused to other proteins. These fusion proteins can be used for a variety of applications. For example, fusion of the present polypeptides to His-tag, HA-tag, protein A, IgG domains, and maltose binding protein facilitates purification. (See Example 5; see also EP A 394,827; Trautnecker, et al., Nature 331:84-86 (1988).) Similarly, fusion to IgG-1, IgG-3, and albumin increases the half-life time in vivo. Nuclear localization signals fused to the

35 polypeptides of the present invention can target the protein to a specific subcellular localization, while covalent heterodimer or homodimers can increase or decrease the activity of a fusion protein. Fusion proteins can also create chimeric molecules having

more than one function. Finally, fusion proteins can increase solubility and/or stability of the fused protein compared to the non-fused protein. All of the types of fusion proteins described above can be made by modifying the following protocol, which outlines the fusion of a polypeptide to an IgG molecule, or the protocol described in Example 5.

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5' and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector.

For example, if pC4 (Accession No.209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 1, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced.

If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. (See, e.g., WO 96/34891.)

Human IgG Fc region:

GGGATCCGGAGGCCAAATCTTCTGACAAAACCTCACATGCCACCGTGCC  
CAGCACCTGAATTCGAGGGTGCACCGTCAGTCTTCTCTTCCCCCAAAACC  
CAAGACACCCCTCATGATCTCCCGACTCTCTGAGGTACATGGTGGTGGT  
GGACGTAGCCACGAGACCCCTGAGGTCAAGTTCAACTGGTACGTGGACG  
GCGTGGAGGTGCATAATGCCAAGACAAAGCCGGGAGGAGCAGTACAAC  
AGCACGTACCGTGTGTGTCAGCGTCTCTCACCGTCTGACCCAGGACTGGCTG  
AATGGCAAGGAGTACAAGTGCAAGGTCTCCAACAAGCCCTCCCAACCCCC  
ATCGAGAAAACCATCTCCAAAGCCAAAGGCGAGCCCGCAGAACCAAGGT  
GTACACCCCTGCCCCCATCCGGGATGAGCTGACCAAGAACAGGTGAGCCT  
GACCTGCCTGGTCAAAGGCTTCTATCCAAGCGACATCGCCGTGGAGTGGA  
GAGCAATGGGACGCCGGAGAACTACAAGACCAAGCCTCCCGTGGTCTGG  
ACTCCGACGGTCTCTTCTCTACAGCAAGCTCACCGTGGACAAAGAGCA  
GGTGGCAGCAGGGGAACGTCTTCTCATGTCCGTGATGCGATGAGGCTCTGC  
ACAACCACTACACGACAGAGCCCTCTCCCTGTCTCCGGGTAAATGAGTGC  
GACGGCCGCACTCTAGAGGAT (SEQ ID NO:1)

### Example 10: Production of an Antibody from a Polypeptide

The antibodies of the present invention can be prepared by a variety of methods. (See, Current Protocols, Chapter 2.) For example, cells expressing a polypeptide of the present invention is administered to an animal to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of the secreted protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology. (Köhler et al., Nature 256:495 (1975); Köhler et al., Eur. J. Immunol. 6:511 (1976); Köhler et al., Eur. J. Immunol. 6:292 (1976); Hammerling et al., in: Monoclonal Antibodies and T-Cell Hybridomas, Elsevier, N.Y., pp. 563-681 (1981).) In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 µg/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100 µg/ml of streptomycin.

The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP2O), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands et al. (Gastroenterology 80:225-232 (1981).) The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide.

Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody

whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein-specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies.

It will be appreciated that Fab and F(ab)<sub>2</sub> and other fragments of the antibodies of the present invention may be used according to the methods disclosed herein. Such fragments are typically produced by proteolytic cleavage, using enzymes such as papain (to produce Fab fragments) or pepsin (to produce F(ab)<sub>2</sub> fragments). Alternatively, secreted protein-binding fragments can be produced through the application of recombinant DNA technology or through synthetic chemistry.

For in vivo use of antibodies in humans, it may be preferable to use "humanized" chimeric monoclonal antibodies. Such antibodies can be produced using genetic constructs derived from hybridoma cells producing the monoclonal antibodies described above. Methods for producing chimeric antibodies are known in the art. (See, for review, Morrison, Science 229:1202 (1985); Oi et al., BioTechniques 4:214 (1986); Cabilly et al., U.S. Patent No. 4,816,567; Taniguchi et al., EP 171496; Morrison et al., EP 173494; Neuberger et al., WO 8601533; Robinson et al., WO 8702671; Boulianne et al., Nature 312:643 (1984); Neuberger et al., Nature 314:268 (1985).)

#### Example 11: Production Of Secreted Protein For High-Throughput Screening Assays

The following protocol produces a supernatant containing a polypeptide to be tested. This supernatant can then be used in the Screening Assays described in Examples 13-20.

First, dilute Poly-D-Lysine (644 587 Boehringer-Mannheim) stock solution (1mg/ml in PBS) 1:20 in PBS (w/o calcium or magnesium 17-516F Biowhitaker) for a working solution of 50ug/ml. Add 200 ul of this solution to each well (24 well plates) and incubate at RT for 20 minutes. Be sure to distribute the solution over each well (note: a 12-channel pipetter may be used with tips on every other channel). Aspirate off the Poly-D-Lysine solution and rinse with 1ml PBS (Phosphate Buffered Saline). The PBS should remain in the well until just prior to plating the cells and plates may be poly-lysine coated in advance for up to two weeks.

Plate 293T cells (do not carry cells past P+20) at 2 x 10<sup>4</sup> cells/well in 5ml DMEM(Dulbecco's Modified Eagle Medium)(with 4.5 G/L glucose and L-glutamine (12-604F Biowhitaker))/10% heat inactivated FBS(14-503F Biowhitaker)/1x Penstrep(17-602E Biowhitaker). Let the cells grow overnight.

The next day, mix together in a sterile solution basin: 300 ul Lipofectamine (18324-012 Gibco/BRL) and 5ml Optimem I (31983070 Gibco/BRL)/96-well plate. With a small volume multi-channel pipetter, aliquot approximately 2ug of an expression

vector containing a polynucleotide insert, produced by the methods described in Examples 8 or 9, into an appropriately labeled 96-well round bottom plate. With a multi-channel pipetter, add 50ul of the Lipofectamine/Optimem I mixture to each well. Pipette up and down gently to mix. Incubate at RT 15-45 minutes. After about 20 minutes, use a multi-channel pipetter to add 150ul Optimem I to each well. As a control, one plate of vector DNA lacking an insert should be transfected with each set of transfections.

Preferably, the transfection should be performed by tag-teaming the following tasks. By tag-teaming, hands on time is cut in half, and the cells do not spend too much time on PBS. First, person A aspirates off the media from four 24-well plates of cells, and then person B rinses each well with .5-1ml PBS. Person A then aspirates off PBS rinse, and person B, using a 12-channel pipetter with tips on every other channel, adds the 200ul of DNA/Lipofectamine/Optimem I complex to the odd wells first, then to the even wells, to each row on the 24-well plates. Incubate at 37°C for 6 hours.

While cells are incubating, prepare appropriate media, either 1%BSA in DMEM with 1x penstrep, or CHO-5 media (see below) with 2mm glutamine and 1x penstrep. (BSA (81-068-3 Bayer) 100gm dissolved in 1L DMEM for a 10% BSA stock solution). Filter the media and collect 50 ul for endotoxin assay in 15ml polystyrene conical.

The transfection reaction is terminated, preferably by tag-teaming, at the end of the incubation period. Person A aspirates off the transfection media, while person B adds 1.5ml appropriate media to each well. Incubate at 37°C for 45 or 72 hours depending on the media used: 1%BSA for 45 hours or CHO-5 for 72 hours.

On day four, using a 300ul multichannel pipetter, aliquot 600ul in one 1ml deep well plate and the remaining supernatant into a 2ml deep well. The supernatants from each well can then be used in the assays described in Examples 13-20.

It is specifically understood that when activity is obtained in any of the assays described below using a supernatant, the activity originates from either the polypeptide directly (e.g., as a secreted protein) or by the polypeptide inducing expression of other proteins, which are then secreted into the supernatant. Thus, the invention further provides a method of identifying the protein in the supernatant characterized by an activity in a particular assay.

**HGS-CHO-5 medium formulation:****Inorganic Salts**

CaCl <sub>2</sub> (anhyd)	116.6 mg/L
CuSO <sub>4</sub> ·5H <sub>2</sub> O	0.00130
Fe(NO <sub>3</sub> ) <sub>3</sub> ·9H <sub>2</sub> O	0.050
FeSO <sub>4</sub> ·7H <sub>2</sub> O	0.417
KCl	311.80
MgCl <sub>2</sub>	28.64
MgSO <sub>4</sub>	48.84
NaCl	6995.50
NaHCO <sub>3</sub>	2400.0
NaH <sub>2</sub> PO <sub>4</sub> ·H <sub>2</sub> O	62.50
Na <sub>2</sub> HPO <sub>4</sub>	71.02
ZnSO <sub>4</sub> ·7H <sub>2</sub> O	4320

**5 Lipids**

Arachidonic Acid	.002 mg/L
Cholesterol	1.022
DL-alpha-Tocopherol-Acetate	.070
Linoleic Acid	0.0520
Linolenic Acid	0.010
Myristic Acid	0.010
Oleic Acid	0.010
Palmitic Acid	0.010
Palmitic Acid	0.010
Pluronic F-68	100
Stearic Acid	0.010
Tween 80	2.20

**Carbon Source**

D-Glucose	4551 mg/L
-----------	-----------

**Amino Acids**

L-Alanine	130.85 mg/ml
L-Arginine-HCL	147.50
L-Asparagine-H <sub>2</sub> O	7.50
L-Aspartic Acid	6.65
L-Cystine-2HCL-H <sub>2</sub> O	29.56
L-Cystine-2HCL	31.29
L-Glutamic Acid	7.35
L-Glutamine	365.0
Glycine	18.75
L-Histidine-HCL-	52.48

H <sub>2</sub> O	
L-Isoleucine	106.97
L-Leucine	111.45
L-Lysine HCL	163.75
L-Methionine	32.34
L-Phenylalanine	68.48
L-Proline	40.0
L-Serine	26.25
L-Threonine	101.05
L-Tryptophan	19.22
L-Tyrosine-2Na-2H <sub>2</sub> O	91.79
L-Valine	99.65

**Vitamins**

Biotin	0.0035 mg/L
D-Ca Pantothenate	3.24
Choline Chloride	11.78
Folic Acid	4.65
I-Inositol	15.60
Niacinamide	3.02
Pyridoxal HCL	3.00
Pyridoxine HCL	0.031
Riboflavin	0.319
Thiamine HCL	3.17
Thymidine	0.365
Vitamin B <sub>12</sub>	0.680

**Other Components**

HEPES Buffer	25 mM
Na Hypoxanthine	2.39 mg/L
Lipoic Acid	0.105
Sodium Putrescine-2HCL	0.084
Sodium Pyruvate	55.0
Sodium Selenite	0.0067
Ethanolamine	20mM
Ferric Citrate	0.122
Methyl-B-Cyclodextrin complexed with Linoleic Acid	41.70
Methyl-B-Cyclodextrin complexed with Oleic Acid	33.33
Methyl-B-Cyclodextrin complexed with Retinal Acetate	10

Adjust osmolarity to 327 mOsm

**Example 12. Construction of GAS Reporter Construct**

One signal transduction pathway involved in the differentiation and proliferation of cells is called the Jaks-STATs pathway. Activated proteins in the Jaks-STATs pathway bind to gamma activation site "GAS" elements or interferon-sensitive responsive element ("ISRE"), located in the promoter of many genes. The binding of a protein to these elements alter the expression of the associated gene.

GAS and ISRE elements are recognized by a class of transcription factors called Signal Transducers and Activators of Transcription, or "STATs." There are six members of the STATs family. Stat1 and Stat3 are present in many cell types, as is Stat2 (as response to IFN-alpha is widespread). Stat4 is more restricted and is not in many cell types though it has been found in T helper class 1 cells after treatment with IL-12. Stat5 was originally called mammary growth factor, but has been found at higher concentrations in other cells including myeloid cells. It can be activated in tissue culture cells by many cytokines.

The STATs are activated to translocate from the cytoplasm to the nucleus upon tyrosine phosphorylation by a set of kinases known as the Janus Kinase ("Jaks") family. Jaks represent a distinct family of soluble tyrosine kinases and include Tyk2, Jak1, Jak2, and Jak3. These kinases display significant sequence similarity and are generally catalytically inactive in resting cells.

The Jaks are activated by a wide range of receptors summarized in the Table below. (Adapted from review by Schidler and Darnell, Ann. Rev. Biochem. 64:621-51 (1995).) A cytokine receptor family, capable of activating Jaks, is divided into two groups: (a) Class 1 includes receptors for IL-2, IL-3, IL-4, IL-6, IL-7, IL-9, IL-11, IL-12, IL-15, Epo, PRL, GH, G-CSF, GM-CSF, LIF, CNTF, and thrombopoietin; and (b) Class 2 includes IFN-a, IFN-g, and IL-10. The Class 1 receptors share a conserved cysteine motif (a set of four conserved cysteines and one tryptophan) and a WSXWS motif (a membrane proximal region encoding Trp-Ser-Xxx-Trp-Ser (SEQ ID NO:2)).

Thus, on binding of a ligand to a receptor, Jaks are activated, which in turn activate STATs, which then translocate and bind to GAS elements. This entire process is encompassed in the Jaks-STATs signal transduction pathway.

Therefore, activation of the Jaks-STATs pathway, reflected by the binding of the GAS or the ISRE element, can be used to indicate proteins involved in the proliferation and differentiation of cells. For example, growth factors and cytokines are known to activate the Jaks-STATs pathway. (See Table below.) Thus, by using GAS elements linked to reporter molecules, activators of the Jaks-STATs pathway can be identified.

	ISRE Ligand	cyk2	JAKs Jak1 Jak2 Jak3	STATs	GAS(elements) or
5	IFN family IFN-a/B IFN-g (IRF1>Lys6>IFP) IL-10	+	+	1,2,3 1 1,3	ISRE GAS
10	gp130 family IL-6 (Pleiotrohic) (IRF1>Lys6>IFP) IL-11(Pleiotrohic) OnM(Pleiotrohic) LIF(Pleiotrohic) CNTF(Pleiotrohic) G-CSF(Pleiotrohic) IL-12(Pleiotrohic)	+	+	1,3 ? ? ? ? ? ? ?	GAS
15	g-c family IL-2 (lymphocytes) IL-4 (lymph/myeloid) >>Ly6(IgH) IL-7 (lymphocytes) IL-9 (lymphocytes) IL-13 (lymphocyte) IL-15	-	-	1,3,5 6 5 5 6 5	GAS GAS (IRF1 = IFP)
20	gp140 family IL-3 (myeloid) (IRF1>IFP>>Ly6) IL-5 (myeloid) GM-CSF (myeloid)	-	+	5 5 5	GAS
25	Growth hormone family GH PRL EPO CAS>IRF1=IFP>>Ly6)	-	+	5 1,3,5 5	GAS(B-
30	Receptor Tyrosine Kinases EGF	+	+	1,3	GAS (IRF1)
35	PDGF CSF-1	?	+	1,3 1,3	GAS (not IRF1)

To construct a synthetic GAS containing promoter element, which is used in the Biological Assays described in Examples 13-14, a PCR based strategy is employed to generate a GAS-SV40 promoter sequence. The 5' primer contains four tandem copies of the GAS binding site found in the IRF1 promoter and previously demonstrated to bind STATs upon induction with a range of cytokines (Rothman et al., Immunity 1:457-468 (1994)), although other GAS or ISRE elements can be used instead. The 5' primer also contains 18bp of sequence complementary to the SV40 early promoter sequence and is flanked with an XhoI site. The sequence of the 5' primer is:

5': GCGCCTCGAGATTCCCGGAAATCTGCGCATCTCAATTAG:3' (SEQ ID NO:3)  
 10 AAATGATTTCCTCCGAAATATCTGCCATCTCAATTAG:3' (SEQ ID NO:4)

The downstream primer is complementary to the SV40 promoter and is flanked with a Hind III site: 5': GCGGCAAGCTTTTGTCAAAGCCTAGGC:3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the B-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI/Hind III and subcloned into BLSK2-. (Stratagene.) Sequencing with forward and reverse primers confirms that the insert contains the following sequence:

5': CTGAGAGATTTCCTCCGAAATCTAGATTTCCTCCGAAATGATTTCCTCCGAAATG  
 20 ATTTCCTCCGAAATATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCC  
 CTAATCTCCGCCCACCTCCGCCCCTAAGTCCGCCAGTTCGCCCATCTCCGC  
 CCATGGCTGACTAATTTTATTATTCAGAGGCCGAGGCCGCTCCGC  
 CTCTGAGCTATTCAGAAAGTAGTGAGGAGGCTTTTTCGAGGCCCTAGGCTTT  
 TGCAAAAGGCTT:3' (SEQ ID NO:5)

With this GAS promoter element linked to the SV40 promoter, a GAS:SEAP2 reporter construct is next engineered. Here, the reporter molecule is a secreted alkaline phosphatase, or "SEAP." Clearly, however, any reporter molecule can be instead of SEAP, in this or in any of the other Examples. Well known reporter molecules that can be used instead of SEAP include chloramphenicol acetyltransferase (CAT), luciferase, alkaline phosphatase, B-galactosidase, green fluorescent protein (GFP), or any protein detectable by an antibody.

The above sequence confirmed synthetic GAS-SV40 promoter element is subcloned into the pSEAP-Promoter vector obtained from Clontech using HindIII and XhoI, effectively replacing the SV40 promoter with the amplified GAS:SV40 promoter element, to create the GAS-SEAP vector. However, this vector does not contain a

neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

Thus, in order to generate mammalian stable cell lines expressing the GAS-SEAP reporter, the GAS-SEAP cassette is removed from the GAS-SEAP vector using SalI and NotI, and inserted into a backbone vector containing the neomycin resistance gene, such as pGFP-1 (Clontech), using these restriction sites in the multiple cloning site, to create the GAS-SEAP/Neo vector. Once this vector is transfected into mammalian cells, this vector can then be used as a reporter molecule for GAS binding as described in Examples 13-14.

Other constructs can be made using the above description and replacing GAS with a different promoter sequence. For example, construction of reporter molecules containing NFK-B and EGR promoter sequences are described in Examples 15 and 16. However, many other promoters can be substituted using the protocols described in these Examples. For instance, SRE, IL-2, NFAT, or Osteocalcin promoters can be substituted, alone or in combination (e.g., GAS/NF-KB/EGR, GAS/NF-KB, IL-2/NFAT, or NF-KB/GAS). Similarly, other cell lines can be used to test reporter construct activity, such as HELA (epithelial), HUVEC (endothelial), Reh (B-cell), Saos-2 (osteoblast), HUVAC (aortic), or Cardiomyocyte.

#### Example 13: High-Throughput Screening Assay for T-cell Activity.

The following protocol is used to assess T-cell activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate T-cells. T-cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The T-cell used in this assay is Jurkat T-cells (ATCC Accession No. TIB-152), although Molt-3 cells (ATCC Accession No. CRL-1552) and Molt-4 cells (ATCC Accession No. CRL-1582) cells can also be used.

Jurkat T-cells are lymphoblastic CD4+ Th1 helper cells. In order to generate stable cell lines, approximately 2 million Jurkat cells are transfected with the GAS-SEAP/Neo vector using DMRIE-C (Life Technologies)(transfection procedure described below). The transfected cells are seeded to a density of approximately 20,000 cells per well and transfectants resistant to 1 mg/ml gentamicin selected. Resistant colonies are expanded and then tested for their response to increasing concentrations of interferon gamma. The dose response of a selected clone is demonstrated.

Specifically, the following protocol will yield sufficient cells for 75 wells containing 200 ul of cells. Thus, it is either scaled up, or performed in multiple to generate sufficient cells for multiple 96 well plates. Jurkat cells are maintained in RPMI

+ 10% serum with 1% Pen-Strep. Combine 2.5 ml of OPTI-MEM (Life Technologies) with 10 ug of plasmid DNA in a T25 flask. Add 2.5 ml OPTI-MEM containing 50 ul of DMRIE-C and incubate at room temperature for 15-45 mins.

During the incubation period, count cell concentration, spin down the required number of cells ( $10^7$  per transfection), and resuspend in OPTI-MEM to a final concentration of  $10^7$  cells/ml. Then add 1 ml of  $1 \times 10^7$  cells in OPTI-MEM to T25 flask and incubate at 37°C for 6 hrs. After the incubation, add 10 ml of RPMI + 15% serum.

The Jurkat:GAS-SEAP stable reporter lines are maintained in RPMI + 10% serum, 1 mg/ml Gentamicin, and 1% Pen-Strep. These cells are treated with supernatants containing a polypeptide as produced by the protocol described in Example 11.

On the day of treatment with the supernatant, the cells should be washed and resuspended in fresh RPMI + 10% serum to a density of 500,000 cells per ml. The exact number of cells required will depend on the number of supernatants being screened. For one 96 well plate, approximately 10 million cells (for 10 plates, 100 million cells) are required.

Transfer the cells to a triangular reservoir boat, in order to dispense the cells into a 96 well dish, using a 12 channel pipette. Using a 12 channel pipette, transfer 200 ul of cells into each well (therefore adding 100,000 cells per well).

After all the plates have been seeded, 50 ul of the supernatants are transferred directly from the 96 well plate containing the supernatants into each well using a 12 channel pipette. In addition, a dose of exogenous interferon gamma (0.1, 1.0, 10 ng) is added to wells H9, H10, and H11 to serve as additional positive controls for the assay.

The 96 well dishes containing Jurkat cells treated with supernatants are placed in an incubator for 48 hrs (note: this time is variable between 48-72 hrs). 35 ul samples from each well are then transferred to an opaque 96 well plate using a 12 channel pipette. The opaque plates should be covered (using sellophene covers) and stored at 20°C until SEAP assays are performed according to Example 17. The plates containing the remaining treated cells are placed at 40°C and serve as a source of material for repeating the assay on a specific well if desired.

As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate Jurkat T cells. Over 30 fold induction is typically observed in the positive control wells.

#### Example 14: High-Throughput Screening Assay Identifying Myeloid Activity

The following protocol is used to assess myeloid activity by identifying factors, such as growth factors and cytokines, that may proliferate or differentiate myeloid cells. Myeloid cell activity is assessed using the GAS/SEAP/Neo construct produced in Example 12. Thus, factors that increase SEAP activity indicate the ability to activate the Jaks-STATS signal transduction pathway. The myeloid cell used in this assay is U937, a pre-monocyte cell line, although TF-1, HL60, or KG1 can be used.

To transiently transfect U937 cells with the GAS/SEAP/Neo construct produced in Example 12, a DEAE-Dextran method (Kharbanda et al., 1994, Cell Growth & Differentiation, 5:259-265) is used. First, harvest  $2 \times 10^6$  U937 cells and wash with PBS. The U937 cells are usually grown in RPMI 1640 medium containing 10% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 mg/ml streptomycin.

Next, suspend the cells in 1 ml of 20 mM Tris-HCl (pH 7.4) buffer containing 0.5 mg/ml DEAE-Dextran, 8 ug GAS-SEAP2 plasmid DNA, 140 mM NaCl, 5 mM KCl, 375 uM  $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$ , 1 mM  $\text{MgCl}_2$ , and 675 uM  $\text{CaCl}_2$ . Incubate at 37°C for 45 min.

Wash the cells with RPMI 1640 medium containing 10% FBS and then resuspend in 10 ml complete medium and incubate at 37°C for 36 hr.

The GAS-SEAP/U937 stable cells are obtained by growing the cells in 400 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 400 ug/ml G418 for couple of passages.

These cells are tested by harvesting  $1 \times 10^6$  cells (this is enough for ten 96-well plates assay) and wash with PBS. Suspend the cells in 200 ml above described growth medium, with a final density of  $5 \times 10^5$  cells/ml. Plate 200 ul cells per well in the 96-well plate (or  $1 \times 10^5$  cells/well).

Add 50 ul of the supernatant prepared by the protocol described in Example 17. Incubate at 37°C for 48 to 72 hr. As a positive control, 100 Unit/ml interferon gamma can be used which is known to activate U937 cells. Over 30 fold induction is typically observed in the positive control wells. SEAP assay the supernatant according to the protocol described in Example 17.

#### **Example 15: High-Throughput Screening Assay Identifying Neuronal Activity.**

When cells undergo differentiation and proliferation, a group of genes are activated through many different signal transduction pathways. One of these genes, EGR1 (early growth response gene 1), is induced in various tissues and cell types upon activation. The promoter of EGR1 is responsible for such induction. Using the EGR1 promoter linked to reporter molecules, activation of cells can be assessed.

Particularly, the following protocol is used to assess neuronal activity in PC12 cell lines. PC12 cells (rat pheochromocytoma cells) are known to proliferate and/or differentiate by activation with a number of mitogens, such as TPA (tetradecanoyl phorbol acetate), NGF (nerve growth factor), and EGF (epidermal growth factor). The EGR1 gene expression is activated during this treatment. Thus, by stably transfecting PC12 cells with a construct containing an EGR promoter linked to SEAP reporter, activation of PC12 cells can be assessed.

The EGR/SEAP reporter construct can be assembled by the following protocol. The EGR-1 promoter sequence (-633 to +1)(Sakamoto K et al., Oncogene 6:867-871 (1991)) can be PCR amplified from human genomic DNA using the following primers:

5' GCGCTCAGGGATGACAGCGATAGAACCCCG-3' (SEQ ID NO:6)  
5' GCGAAGCTTCGCGACTCCCGGATCCGCCTC-3' (SEQ ID NO:7)

Using the GAS:SEAP/Neo vector produced in Example 12, EGR1 amplified product can then be inserted into this vector. Linearize the GAS:SEAP/Neo vector using restriction enzymes XhoI/HindIII, removing the GAS/SV40 stuffer. Restrict the EGR1 amplified product with these same enzymes. Ligate the vector and the EGR1 promoter.

To prepare 96 well-plates for cell culture, two mls of a coating solution (1:30 dilution of collagen type I (Upstate Biotech Inc. Cat#08-115) in 30% ethanol (filter sterilized)) is added per one 10 cm plate or 50 ml per well of the 96-well plate, and allowed to air dry for 2 hr.

PC12 cells are routinely grown in RPMI-1640 medium (Bio Whittaker) containing 10% horse serum (JRH BIOSCIENCES, Cat. # 12449-78P), 5% heat-inactivated fetal bovine serum (FBS) supplemented with 100 units/ml penicillin and 100 ug/ml streptomycin on a precoated 10 cm tissue culture dish. One to four split is done every three to four days. Cells are removed from the plates by scraping and resuspended with pipetting up and down for more than 15 times.

Transfect the EGR/SEAP/Neo construct into PC12 using the Lipofectamine protocol described in Example 11. EGR-SEAP/PC12 stable cells are obtained by growing the cells in 300 ug/ml G418. The G418-free medium is used for routine growth but every one to two months, the cells should be re-grown in 300 ug/ml G418 for couple of passages.

To assay for neuronal activity, a 10 cm plate with cells around 70 to 80% confluent is screened by removing the old medium. Wash the cells once with PBS (Phosphate buffered saline). Then starve the cells in low serum medium (RPMI-1640 containing 1% horse serum and 0.5% FBS with antibiotics) overnight.

The next morning, remove the medium and wash the cells with PBS. Scrape off the cells from the plate, suspend the cells well in 2 ml low serum medium. Count the cell number and add more low serum medium to reach final cell density as  $5 \times 10^5$  cells/ml.

Add 200 ul of the cell suspension to each well of 96-well plate (equivalent to  $1 \times 10^5$  cells/well). Add 50 ul supernatant produced by Example 11, 37°C for 48 to 72 hr. As a positive control, a growth factor known to activate PC12 cells through EGR can be used, such as 50 ng/ul of Neuronal Growth Factor (NGF). Over fifty-fold induction of SEAP is typically seen in the positive control wells. SEAP assay the supernatant according to Example 17.

#### **Example 16: High-Throughput Screening Assay for T-cell Activity**

NF- $\kappa$ B (Nuclear Factor  $\kappa$ B) is a transcription factor activated by a wide variety of agents including the inflammatory cytokines IL-1 and TNF, CD30 and CD40, lymphotoxin-alpha and lymphotoxin-beta, by exposure to LPS or thrombin, and by expression of certain viral gene products. As a transcription factor, NF- $\kappa$ B regulates the expression of genes involved in immune cell activation, control of apoptosis (NF- $\kappa$ B appears to shield cells from apoptosis), B and T-cell development, anti-viral and antimicrobial responses, and multiple stress responses.

In non-stimulated conditions, NF- $\kappa$ B is retained in the cytoplasm with I- $\kappa$ B (Inhibitor  $\kappa$ B). However, upon stimulation, I- $\kappa$ B is phosphorylated and degraded, causing NF- $\kappa$ B to shuttle to the nucleus, thereby activating transcription of target genes. Target genes activated by NF- $\kappa$ B include IL-2, IL-6, GM-CSF, ICAM-1 and class I MHC.



Due to its central role and ability to respond to a range of stimuli, reporter constructs utilizing the NF- $\kappa$ B promoter element are used to screen the supernatants produced in Example 11. Activators or inhibitors of NF- $\kappa$ B would be useful in treating diseases. For example, inhibitors of NF- $\kappa$ B could be used to treat those diseases related to the acute or chronic activation of NF- $\kappa$ B, such as rheumatoid arthritis.

To construct a vector containing the NF- $\kappa$ B promoter element, a PCR based

strategy is employed. The upstream primer contains four tandem copies of the NF- $\kappa$ B binding site (GGGGACTTCCC) (SEQ ID NO:8), 18 bp of sequence complementary to the 5' end of the SV40-early promoter sequence, and is flanked with an XhoI site:

5'-GCGGCCCTCGAGGGGACTTCCCGGGGACTTCCCGGGGACTTCCCGGGGAC  
TTTCCATCTGCCATCTCAATTAG-3' (SEQ ID NO:9)

The downstream primer is complementary to the 3' end of the SV40 promoter and is flanked with a Hind III site:

5'-GCGGCAAGCTTTTGCAGAGCCTAGGC-3' (SEQ ID NO:4)

PCR amplification is performed using the SV40 promoter template present in the pB-gal:promoter plasmid obtained from Clontech. The resulting PCR fragment is digested with XhoI and Hind III and subcloned into BLSK2<sup>-</sup>. (Stratagene) Sequencing with the T7 and T3 primers confirms the insert contains the following sequence:

20 5'-CTCGAGGGGACTTCCCGGGGACTTCCCGGGGACTTCCCGGGGACTTTC  
ATCTGCCATCTCAATTAGTCAGCAACCATAGTCCCGCCCTAATCCGCCCA  
TCCCGCCCTAATCCGCCCAAGTTCGCCCAATTCTCCGCCCATGGCTGACT  
AATTTTATTATGAGAGGCGGCGCTCGGCTCTGAGCTATTC  
CAGAAAGTAGGAGGCTTTTGGAGGCGCTAGGCTTTTGCAAAAAGCTT:  
25 3' (SEQ ID NO:10)

Next, replace the SV40 minimal promoter element present in the pSEAP2-

promoter plasmid (Clontech) with this NF- $\kappa$ B/SV40 fragment using XhoI and HindIII. However, this vector does not contain a neomycin resistance gene, and therefore, is not preferred for mammalian expression systems.

In order to generate stable mammalian cell lines, the NF- $\kappa$ B/SV40/SEAP cassette is removed from the above NF- $\kappa$ B/SEAP vector using restriction enzymes SalI and NotI, and inserted into a vector containing neomycin resistance. Particularly, the

NF- $\kappa$ B/SV40/SEAP cassette was inserted into pGFP-1 (Clontech), replacing the GFP gene, after restricting pGFP-1 with SalI and NotI.

Once NF- $\kappa$ B/SV40/SEAP/Neo vector is created, stable Jurkat T-cells are created and maintained according to the protocol described in Example 13. Similarly, the method for assaying supernatants with these stable Jurkat T-cells is also described in Example 13. As a positive control, exogenous TNF alpha (0.1, 1, 10 ng) is added to wells H9, H10, and H11, with a 5-10 fold activation typically observed.

#### Example 17: Assay for SEAP Activity

As a reporter molecule for the assays described in Examples 13-16, SEAP activity is assayed using the Tropix Phospho-light Kit (Cat. BP-400) according to the following general procedure. The Tropix Phospho-light Kit supplies the Dilution, Assay, and Reaction Buffers used below.

Prime a dispenser with the 2.5x Dilution Buffer and dispense 15  $\mu$ l of 2.5x dilution buffer into Optiplates containing 35  $\mu$ l of a supernatant. Seal the plates with a plastic sealer and incubate at 65°C for 30 min. Separate the Optiplates to avoid uneven heating.

Cool the samples to room temperature for 15 minutes. Empty the dispenser and prime with the Assay Buffer. Add 50  $\mu$ l Assay Buffer and incubate at room

temperature 5 min. Empty the dispenser and prime with the Reaction Buffer (see the table below). Add 50  $\mu$ l Reaction Buffer and incubate at room temperature for 20 minutes. Since the intensity of the chemiluminescent signal is time dependent, and it takes about 10 minutes to read 5 plates on luminometer, one should treat 5 plates at each time and start the second set 10 minutes later.

Read the relative light unit in the luminometer. Set H12 as blank, and print the results. An increase in chemiluminescence indicates reporter activity.

#### Reaction Buffer Formulation:

# of plates	Rxn buffer diluent (ml)	CSPD (ml)
10	60	3
11	65	3.25
12	70	3.5
13	75	3.75
14	80	4

15	85	4.25
16	90	4.5
17	95	4.75
18	100	5
19	105	5.25
20	110	5.5
21	115	5.75
22	120	6
23	125	6.25
24	130	6.5
25	135	6.75
26	140	7
27	145	7.25
28	150	7.5
29	155	7.75
30	160	8
31	165	8.25
32	170	8.5
33	175	8.75
34	180	9
35	185	9.25
36	190	9.5
37	195	9.75
38	200	10
39	205	10.25
40	210	10.5
41	215	10.75
42	220	11
43	225	11.25
44	230	11.5
45	235	11.75
46	240	12
47	245	12.25
48	250	12.5
49	255	12.75
50	260	13

**Example 18: High-Throughput Screening Assay Identifying Changes in Small Molecule Concentration and Membrane Permeability**

Binding of a ligand to a receptor is known to alter intracellular levels of small molecules, such as calcium, potassium, sodium, and pH, as well as alter membrane potential. These alterations can be measured in an assay to identify supernatants which bind to receptors of a particular cell. Although the following protocol describes an assay for calcium, this protocol can easily be modified to detect changes in potassium, sodium, pH, membrane potential, or any other small molecule which is detectable by a fluorescent probe.

The following assay uses Fluorometric Imaging Plate Reader ("FLIPR") to measure changes in fluorescent molecules (Molecular Probes) that bind small molecules. Clearly, any fluorescent molecule detecting a small molecule can be used instead of the calcium fluorescent molecule, fluo-3, used here.

For adherent cells, seed the cells at 10,000-20,000 cells/well in a Co-star black 96-well plate with clear bottom. The plate is incubated in a CO<sub>2</sub> incubator for 20 hours. The adherent cells are washed two times in Biotek washer with 200 ul of HBSS (Hank's Balanced Salt Solution) leaving 100 ul of buffer after the final wash.

A stock solution of 1 mg/ml fluo-3 is made in 10% pluronic acid DMSO. To load the cells with fluo-3, 50 ul of 12 ug/ml fluo-3 is added to each well. The plate is incubated at 37°C in a CO<sub>2</sub> incubator for 60 min. The plate is washed four times in the Biotek washer with HBSS leaving 100 ul of buffer.

For non-adherent cells, the cells are spun down from culture media. Cells are re-suspended to 2.5x10<sup>6</sup> cells/ml with HBSS in a 50-ml conical tube. 4 ul of 1 mg/ml fluo-3 solution in 10% pluronic acid DMSO is added to each ml of cell suspension.

The tube is then placed in a 37°C water bath for 30-60 min. The cells are washed twice with HBSS, resuspended to 1x10<sup>6</sup> cells/ml, and dispensed into a microplate, 100 ul/well. The plate is centrifuged at 1000 rpm for 5 min. The plate is then washed once in Denley CellWash with 200 ul, followed by an aspiration step to 100 ul final volume.

For a non-cell based assay, each well contains a fluorescent molecule, such as fluo-3. The supernatant is added to the well, and a change in fluorescence is detected.

To measure the fluorescence of intracellular calcium, the FLIPR is set for the following parameters: (1) System gain is 300-800 mW; (2) Exposure time is 0.4 second; (3) Camera F/stop is F/2; (4) Excitation is 488 nm; (5) Emission is 530 nm; and (6) Sample addition is 50 ul. Increased emission at 530 nm indicates an extracellular

signaling even which has resulted in an increase in the intracellular  $Ca^{++}$  concentration.

#### Example 19: High-Throughput Screening Assay Identifying Tyrosine Kinase Activity

The Protein Tyrosine Kinases (PTK) represent a diverse group of transmembrane and cytoplasmic kinases. Within the Receptor Protein Tyrosine Kinase (RPTK) group are receptors for a range of mitogenic and metabolic growth factors including the PDGF, FGF, EGF, NGF, HGF and Insulin receptor subfamilies. In addition there are a large family of RPTKs for which the corresponding ligand is unknown. Ligands for RPTKs include mainly secreted small proteins, but also membrane-bound and extracellular matrix proteins.

Activation of RPTK by ligands involves ligand-mediated receptor dimerization, resulting in transphosphorylation of the receptor subunits and activation of the cytoplasmic tyrosine kinases. The cytoplasmic tyrosine kinases include receptor associated tyrosine kinases of the src-family (e.g., src, yes, lck, lyn, fyn) and non-receptor linked and cytosolic protein tyrosine kinases, such as the Jak family, members of which mediate signal transduction triggered by the cytokine superfamily of receptors (e.g., the Interleukins, Interferons, GM-CSF, and Leptin).

Because of the wide range of known factors capable of stimulating tyrosine kinase activity, the identification of novel human secreted proteins capable of activating tyrosine kinase signal transduction pathways are of interest. Therefore, the following protocol is designed to identify those novel human secreted proteins capable of activating the tyrosine kinase signal transduction pathways.

Seed target cells (e.g., primary keratinocytes) at a density of approximately 25,000 cells per well in a 96 well Loprodyne Silent Screen Plates purchased from Nalge Nunc (Naperville, IL). The plates are sterilized with two 30 minute rinses with 100% ethanol, rinsed with water and dried overnight. Some plates are coated for 2 hr with 100 ml of cell culture grade type I collagen (50 mg/ml), gelatin (2%) or polylysine (50 mg/ml), all of which can be purchased from Sigma Chemicals (St. Louis, MO) or 10% Matrigel purchased from Becton Dickinson (Bedford, MA), or calf serum, rinsed with PBS and stored at 4°C. Cell growth on these plates is assayed by seeding 5,000 cells/well in growth medium and indirect quantitation of cell number through use of alamarBlue as described by the manufacturer Alamar Biosciences, Inc. (Sacramento, CA) after 48 hr. Falcon plate covers #3071 from Becton Dickinson (Bedford, MA) are

used to cover the Loprodyne Silent Screen Plates. Falcon Microtest III cell culture plates can also be used in some proliferation experiments.

To prepare extracts, A431 cells are seeded onto the nylon membranes of

Loprodyne plates (20,000/200ml/well) and cultured overnight in complete medium.

- 5 Cells are quiesced by incubation in serum-free basal medium for 24 hr. After 5-20 minutes treatment with EGF (60ng/ml) or 50 ul of the supernatant produced in Example 11, the medium was removed and 100 ml of extraction buffer (20 mM HEPES pH 7.5, 0.15 M NaCl, 1% Triton X-100, 0.1% SDS, 2 mM Na3VO4, 2 mM Na4P2O7 and a cocktail of protease inhibitors (# 1836170) obtained from Boehringer Mannheim (Indianapolis, IN) is added to each well and the plate is shaken on a rotating shaker for 5 minutes at 40°C. The plate is then placed in a vacuum transfer manifold and the extracts filtered through the 0.45 mm membrane bottoms of each well using house vacuum. Extracts are collected in a 96-well catch/assay plate in the bottom of the vacuum manifold and immediately placed on ice. To obtain extracts clarified by centrifugation, the content of each well, after detergent solubilization for 5 minutes, is removed and centrifuged for 15 minutes at 40°C at 16,000 x g.

Test the filtered extracts for levels of tyrosine kinase activity. Although many methods of detecting tyrosine kinase activity are known, one method is described here.

- 15 Generally, the tyrosine kinase activity of a supernatant is evaluated by determining its ability to phosphorylate a tyrosine residue on a specific substrate (a biotinylated peptide). Biotinylated peptides that can be used for this purpose include PSK1 (corresponding to amino acids 6-20 of the cell division kinase cdc2-p34) and PSK2 (corresponding to amino acids 1-17 of gastrin). Both peptides are substrates for a range of tyrosine kinases and are available from Boehringer Mannheim.

- 25 The tyrosine kinase reaction is set up by adding the following components in order. First, add 10ul of 5uM Biotinylated Peptide, then 10ul ATP/Mg2+ (5mM ATP/50mM MgCl2), then 10ul of 5x Assay Buffer (40mM imidazole hydrochloride, pH7.3, 40 mM beta-glycerophosphate, 1mM EGTA, 100mM MgCl2, 5 mM MnCl2, 0.5 mg/ml BSA), then 5ul of Sodium Vanadate(1mM), and then 5ul of water. Mix the components gently and preincubate the reaction mix at 300C for 2 min. Initial the reaction by adding 10ul of the control enzyme or the filtered supernatant.

The tyrosine kinase assay reaction is then terminated by adding 10 ul of 120mM EDTA and place the reactions on ice.

- 35 Tyrosine kinase activity is determined by transferring 50 ul aliquot of reaction mixture to a microtiter plate (MTP) module and incubating at 370C for 20 min. This

allows the streptavidin coated 96 well plate to associate with the biotinylated peptide. Wash the MTP module with 300ul/well of PBS four times. Next add 75 ul of anti-phosphotyrosine antibody conjugated to horse radish peroxidase(anti-P-Tyr-POD(0.5u/ml)) to each well and incubate at 37°C for one hour. Wash the well as above.

Next add 100ul of peroxidase substrate solution (Boehringer Mannheim) and incubate at room temperature for at least 5 mins (up to 30 min). Measure the absorbance of the sample at 405 nm by using ELISA reader. The level of bound peroxidase activity is quantitated using an ELISA reader and reflects the level of tyrosine kinase activity.

#### **Example 20: High-Throughput Screening Assay Identifying Phosphorylation Activity**

As a potential alternative and/or complement to the assay of protein tyrosine kinase activity described in Example 19, an assay which detects activation (phosphorylation) of major intracellular signal transduction intermediates can also be used. For example, as described below one particular assay can detect tyrosine phosphorylation of the Erk-1 and Erk-2 kinases. However, phosphorylation of other molecules, such as Raf, JNK, p38 MAP, Map kinase kinase (MEK), MEK kinase, Src, Muscle specific kinase (MuSK), IRAK, Tec, and Janus, as well as any other substituting these molecules for Erk-1 or Erk-2 in the following assay.

Specifically, assay plates are made by coating the wells of a 96-well ELISA plate with 0.1ml of protein G (1ug/ml) for 2 hr at room temp. (RT). The plates are then rinsed with PBS and blocked with 3% BSA/PBS for 1 hr at RT. The protein G plates are then treated with 2 commercial monoclonal antibodies (100ng/well) against Erk-1 and Erk-2 (1 hr at RT) (Santa Cruz Biotechnology). (To detect other molecules, this step can easily be modified by substituting a monoclonal antibody detecting any of the above described molecules.) After 3-5 rinses with PBS, the plates are stored at 4°C until use.

A431 cells are seeded at 20,000/well in a 96-well Loprodyne filterplate and cultured overnight in growth medium. The cells are then starved for 48 hr in basal medium (DMEM) and then treated with EGF (6ng/well) or 50 ul of the supernatants obtained in Example 11 for 5-20 minutes. The cells are then solubilized and extracts filtered directly into the assay plate.

After incubation with the extract for 1 hr at RT, the wells are again rinsed. As a positive control, a commercial preparation of MAP kinase (10ng/well) is used in place of A431 extract. Plates are then treated with a commercial polyclonal (rabbit) antibody (1ug/ml) which specifically recognizes the phosphorylated epitope of the Erk-1 and Erk-2 kinases (1 hr at RT). This antibody is biotinylated by standard procedures. The bound polyclonal antibody is then quantitated by successive incubations with Europium-streptavidin and Europium fluorescence enhancing reagent in the Wallac DELFIA instrument (time-resolved fluorescence). An increased fluorescent signal over background indicates a phosphorylation.

#### **Example 21: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide**

RNA isolated from entire families or individual patients presenting with a phenotype of interest (such as a disease) is isolated. cDNA is then generated from these RNA samples using protocols known in the art. (See, Sambrook.) The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO:X. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky, D., et al., Science 252:706 (1991).

PCR products is then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epizentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations is then cloned and sequenced to validate the results of the direct sequencing.

PCR products is cloned into T-tailed vectors as described in Holton, T.A. and Graham, M.W., Nucleic Acids Research, 19:1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals is identified by mutations not present in unaffected individuals.

Genomic rearrangements are also observed as a method of determining alterations in a gene corresponding to a polynucleotide. Genomic clones isolated according to Example 2 are nick-translated with digoxigenindeoxy-uridine 5'-triphosphate (Boehringer Mannheim), and FISH performed as described in Johnson, Cg. et al., Methods Cell Biol. 35:73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human col-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C- and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. (Johnson, C.V. et al., Genet. Anal. Tech. Appl., 8:75 (1991).) Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

#### **Example 22: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample**

A polypeptide of the present invention can be detected in a biological sample, and if an increased or decreased level of the polypeptide is detected, this polypeptide is a marker for a particular phenotype. Methods of detection are numerous, and thus, it is understood that one skilled in the art can modify the following assay to fit their particular needs.

For example, antibody-sandwich ELISAs are used to detect soluble polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 ug/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described in Example 10. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced.

The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide.

Next, 50 ul of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate.

Add 75 ul of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl phosphate (NPP) substrate solution to each well and incubate 1 hour at room temperature. Measure the reaction by a microtiter plate reader. Prepare a standard curve, using serial dilutions of a control sample, and plot polypeptide concentration on

the X-axis (log scale) and fluorescence or absorbance of the Y-axis (linear scale). Interpolate the concentration of the polypeptide in the sample using the standard curve.

#### **Example 23: Formulating a Polypeptide**

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmacologically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 µg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 µg/kg/hour to about 50 µg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracisternally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), buccally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semi-permeable polymer matrices in the form of shaped articles, e.g., films, or microcapsules.

Sustained-release matrices include polylactides (U.S. Pat. No. 3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22:547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15:167-277 (1981), and R. Langer, Chem. Tech. 12:98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric

acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE 3,218,121; Epstein et al., Proc. Natl. Acad. Sci. USA 82:3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77:4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U.S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, i.e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation. For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides.

Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate, succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e.g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, mannose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of

about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile.

Sterility is readily accomplished by filtration through sterile filtration membranes (e.g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampoules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1% (w/v) aqueous polypeptide solution, and resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

#### **Example 24: Method of Treating Decreased Levels of the Polypeptide**

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form. Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 ug/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided in Example 23.

**Example 25: Method of Treating Increased Levels of the Polypeptide**

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided in Example 23.

**Example 26: Method of Treatment Using Gene Therapy**

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e.g., Ham's F12 media, with 10% FBS, penicillin and streptomycin, is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks.

pMV-7 (Kirschmeier, P.T. et al., DNA, 7:219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5' and 3' end sequences respectively as set forth in Example 1. Preferably, the 5' primer contains an EcoRI site and the 3' primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to

transform bacteria HB101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+am12 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media. If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is being produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

It will be clear that the invention may be practiced otherwise than as particularly described in the foregoing description and examples. Numerous modifications and variations of the present invention are possible in light of the above teachings and, therefore, are within the scope of the appended claims.

The entire disclosure of each document cited (including patents, patent applications, journal articles, abstracts, laboratory manuals, books, or other disclosures) in the Background of the Invention, Detailed Description, and Examples is hereby incorporated herein by reference.

## (1) GENERAL INFORMATION:

- (i) APPLICANT: Human Genome Sciences, Inc. et al.  
(ii) TITLE OF INVENTION: 186 Human Secreted Proteins  
(iii) NUMBER OF SEQUENCES: 644  
(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Human Genome Sciences, Inc.  
(B) STREET: 9410 Key West Avenue  
(C) CITY: Rockville  
(D) STATE: Maryland  
(E) COUNTRY: USA  
(F) ZIP: 20850

## (v) COMPUTER READABLE FORM:

- (A) MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage  
(B) COMPUTER: HP Vectra 486/33  
(C) OPERATING SYSTEM: MSDOS version 6.2  
(D) SOFTWARE: ASCII Text

## (vi) CURRENT APPLICATION DATA:

- (A) APPLICATION NUMBER:  
(B) FILING DATE: March 6, 1998  
(C) CLASSIFICATION:

## (vii) PRIOR APPLICATION DATA:

- (A) APPLICATION NUMBER:  
(B) FILING DATE:

## (viii) ATTORNEY/AGENT INFORMATION:

- (A) NAME: A. Anders Brookes, Esq.  
(B) REGISTRATION NUMBER: 36,373  
(C) REFERENCE/DOCKET NUMBER: PS002.PCT

## (vi) TELECOMMUNICATION INFORMATION:

- (A) TELEPHONE: (301) 309-8504  
(B) TELEFAX: (301) 309-8439

## (2) INFORMATION FOR SEQ ID NO: 1:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 733 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GGGATCCGGA GCGCAATCT TCTGACAAA CTCACACATG CCGACGCTGC CCGACGCTGC 60  
AATTGAGGG TCGACCTCTA GTCTTCTCT TCGCGCCAAA ACCGAGGAC ACCCTGATGA 120  
TCTCCCGAC TCTGAGGTC ACATGCGTGG TGGTGGAGCT AAGCCACGAA GACCTCTGAG 180  
TCAGTTTCAA CTGCTAGCTG GACGGCGTGG AGCTGCATTA TCGCAGACAA AAGCGCGCGG 240  
AGGACAGTA CAACGACAG TACCGTGTGG TCAGCGTCTT CACCGTCTCT CACGAGGACT 300  
GGCTGATGG CAGGAGTAC AGTGGCAGG TCTCCACGAA AGGCTTCCCA ACCCGCATGG 360  
AGAAACCAT CTCGAGGCC AAGGGGAGG CCGGAGACC ACAGTGTATC ACCCTGCCCC 420  
CATCCCGGA TCGCTGACC AGAGCCAGG TCAGCGTGCAC CTGCTGCTTC AAGGCTTTCT 480  
ATCCAGGGA CATGCCCTG GAGTGGGAGA GCAATGGGCA GCGCGAGAC AACTACAGCA 540  
CCAGCGCTCC CCGCTGGAC TCCAGCGCT CTCTTCTCT CTACAGCAG CTGACCGTGG 600  
ACAGAGCAG GTGCGAGCAG GCGAGCTCT TCTCATGCTC CTTGATGAT GAGGCTCTGC 660  
ACAGCAGTA CACGAGCAG AGCTCTCCC TGTGTCGCGG TAAATGATG GAGAGCGCGC 720  
GACTCTAGAG GAT 733



## (2) INFORMATION FOR SEQ ID NO: 2:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 5 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Trp Ser Xaa Trp Ser  
1 5

- (2) INFORMATION FOR SEQ ID NO: 3:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 86 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGCGCTGGAG ATTTCGCCGA AATCTAGATT TCCCGGAAT GATTTCGCC GAATGATTTC  
60  
CCGGAATAT CTGCATCTC AATAG  
86

## (2) INFORMATION FOR SEQ ID NO: 4:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

GGCGAAGCT TTTTGCNAAG CCTAGGC  
27

## (2) INFORMATION FOR SEQ ID NO: 5:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 271 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CTCGAGATT CCGGAATC TAGATTCCC CGAATGATT TCCCGGAAT GATTTCGCC  
60

AATATCTGC CACTCATTT AGTCAGCAC CATATCTCCG CCTTAATC CCCCCATCC  
120  
GCCCTAATC CCGCCAGTT CCGCCCATC TCGCCCATC GCGTACTAA TTTTITTTT  
180  
5 TTATCCAGAG CCGCAGCGCG CTTCCGCTC TAGCTATTG CAGAACTACT GAGGAGGCTT  
240  
TTTTCAGGC CTAGCTTTT GCAAAAGCT T  
271

## (2) INFORMATION FOR SEQ ID NO: 6:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CGCTCCAGG GATACAGCG ATAGACCC GG  
32

## (2) INFORMATION FOR SEQ ID NO: 7:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 31 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGGAGCTTC GCGATCCCC GCAATCGCT C  
31

## (2) INFORMATION FOR SEQ ID NO: 8:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 12 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

GGGAGCTTC CC  
12

## (2) INFORMATION FOR SEQ ID NO: 9:

- (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 73 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

GGGCGCTCGA GGGGACTTTC CCGGGACTT TCCGGGACT TCCGGGACT TCCGACTCG 60

10 CCATCGAAT TAG 73

15 (2) INFORMATION FOR SEQ ID NO: 10:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

25 CTCCAGGGGA CTTTCCGGG GACTTCCGG GACTTCCGG GACTTCCGA TCTGCATCT 60  
CAATTATCA GCAACCATAG TCCCGCCCT AACTCCGCC ATCCCGCCC TAACTCCGCC 120  
CAGTTCCGCC CATTTCGCC CCGATGCGT ACTAATTTT TTTATTATG CAGAGCGGA 180  
GGCGCCCTCG GCGCTCAGC TATTCGAAA GTAGTAGGA GCGTTTTCG CAGCGCTAG 240  
CTTTTCGAA AAGCTT 256

35 (2) INFORMATION FOR SEQ ID NO: 11:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 582 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

45 GCGACGAGT AATTCTACC AGAATTCC AGACATTA GTAGTAGAA AAAAAAGAA 60  
50 GCAAGCTGT AAGATCTTG GATCCCTTA TATAGTACT ATAGCTGAA TCTGTATTC 120  
AATCACTTT TCTCTTTAT CCTCAACCA AAAATTGTT TAATTTTGA TCCCAATGT 180  
55 TTTTATCTT TGTATATTT TTAATAATC TTTTCTCTC ATCATGCT TTTTGTGTT 240  
TGTAAATGA CTATCTGCA CTTTGAGAT GATTTACTC TTGTCACTT ACAAATATG 300  
GATATGATA TTTTCTATC AGATGTAGT TTTGACCAA GATTTGCTA TTTGTTTATA 360  
60 AGAAAAAAC TGGCTTCATT TCTGTGAAT TCGCTTTCA AATTTCCTT TTACAGCTGT 420

AAGCAACTG AGATACCTG ATGCTGTGA TTCTTTTCA TGAATCTAC CACTATTTT 480

5 AAGCACTGAG CTTTATATA TTGTCTAAT TGTAAAGTT ATTGTCTTA ACTCATTTA 540

TAATATPACT GTTATCTGT TTCTGAAAA AAAAAAAA AA 582

10 (2) INFORMATION FOR SEQ ID NO: 12:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

20 GTTTGGGGGT GAGGCGGAG TCGTGGGGG CTTCTGCC GCGCAGGACA CAGCTACTG 60  
CAGCGCGCG GCGCTGGCT ATGATGTTCC TACCCAGGG CCGGCTCTG CCTCTACTC 120  
25 GTCCAGGCC CACTTCCAG CAGAGAGCC TCCCAAGCC TTCAGGCGT CTGGAGTCA 180  
CCTTTGGA TGGACTTAAA GGAACCTTGT GTGGAGAG GTGCTGCCA AAGAGCTGC 240  
TCTGTGCTC CAGGAGGCC CTTGTAAG GAGGCGCAG GACTCATCAG GAGCTCCCTG 300  
30 GACCCCTCA GCGCAGGCG CTTGGCGCG ACCCCAGCA TTTGCTCTG CTGCCCCAA 360  
GGGACAGGA AGCTCTTGG GCTCTTCCC TTCTGAGCA AGGCGCGCT CTTTGGCTC 420  
35 AATTAAGTG TACATATTT TCAITTAAG CCTCTTCTT AAAAA 465

40 (2) INFORMATION FOR SEQ ID NO: 13:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 474 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

50 ATGCATTC TCGTCAAGC CTTTCTTTG GTGCCACTC TCGCTCTTG TGAATGCCC 60  
ATATCGTAG CCTTCTCCC CTCGAGAG GCTTCTTGA TAGATTAGA AATAGATG 120  
55 AGTGCAATT CCTATGTGA TATAGAGG ACCCAGAGA CATCTCTTT AATTAAGG 180  
AAGTGTCCA TCCCTTTAG TCGGAATAG AACTTGTGC TCAATCTCT GAGACTAGGC 240  
CTTTAAGCA GCTTCTGTG TTCTCATTT TCCAGTGT TTCGAGGCT TTTATCGAA 300  
60 AGTATGTT CCGTTTAAA TATTTCTTA TGAAGCGCG CTTGTGTGCT CAGCGCTGTA 360

5 ACCTGACAG TTGGGGCTG ACCGGTGA TCACGAGTC AGCAGATCA GACCATCTG 420  
GSTMATCG TGAACCCCG TCTCTACTAA AATACAGAA AAAAAAAAA AAA 474

10 (2) INFORMATION FOR SEQ ID NO: 14:

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 314 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

20 TTAATTTGG GAGCAGACC TGAATGCGG CTTTACATG GAGATATAT TCTGTCTCC 60  
ATCTCAATG CCCAGTACC TGAACGAGA TGAATATAC CACACACAT GTCTCTTAT 120  
CATGATGCG TTATGAGTA GATTATTTT ATCATTGCA TTATTAGCTC TACAGTGTT 180  
TATATATAT TCTCATTTT TAACTCTTC CTTAGTGGC TTTTGTATC AATCTATG 240  
CTCTCAGG CATTGAGCT CTGATCTC CTTATGCG GAGAGCTG TTGAGAGAG 300  
AGATATATAC TTCC 314

## (2) INFORMATION FOR SEQ ID NO: 15:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 613 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

45 CTGATATCG CTTCTGGCTA AAGTGACCA TCCCATGAT CATCTGCTT TTATTATAT 60  
ATGTTCTTAA TGGTGCGAG CAGACAGA AATGAGAGA AAGATGGGT AGCTCAGCA 120  
ACCCACTAAA TCTATCTAT GGGTGGGTT CACCAGGCT GCTTTTGGG TTATGTCTCA 180  
CTATACAGA GCTCCAGG AGACTGGA CTCAGTCC TTAAGCTGCT TACTATAGC 240  
CTACTCTTC GATTACACC AACATATTC CAGCTCATC CTCCTTCCR AAGTGGCTT 300  
TCTGGCCGAG ATCGGATG CATTTACTA ATCTCATAT GAAATGTCCA CAGGATTC 360  
CATTTTATAT AGCATACAT AGTTTGTG CAATTTGCT TTGAGAGAC TCCCATGCA 420  
GCTGCTAGA GACGTATAG GAGGGGCTC TTGAGGCTTT CCGATAGCT TTGAGTGA 480  
60 ATAGCTCTTA GCGAATCG CAGGAGCTC CTGCGCACT GTATTACTG GGAACACTG 540

ATTGCTTACA AATTATCTT CTTTAACTT TTCTAGTTC TTACATTTA CTGCTGAAC 600  
CCAAATATCG CAC 613

## (2) INFORMATION FOR SEQ ID NO: 16:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 356 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

20 CCCCCCAT TGAACCTCG GCTGTGAAG TTTTGGCTG TGTGGTCTT TCTGTGTGC 60  
GCTGTGTG TGAATCCCA CTCCTTTGC AAGTGGGAG CAGCCATCA TGAAGCCGCC 120  
TTATTTTGA TTGAGATGA CCAATCTCC CCCCCAGCG CTCGTCTGCG TCCCTCATG 180  
25 GTGAGTGTG TCCCTGCCA AGAGGCTGA TTGTGGGGA ATGGCATCAT CTATATGAT 240  
GGAGGCAAT TTGTCTCTG TTAATTTAT TACACATCA TTGCACTCTG GCACTCCAT 300  
CCTGGAAC CTATTTTGT GTTTACCA AGGACGAGG GGGAAAAAAAA AAAA 356

## (2) INFORMATION FOR SEQ ID NO: 17:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 414 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 17:

45 GAACTAAT CCCCCGCTT TTACNGTGA CTTGGGAAT AAGTATGCG TAATCTACTAA 60  
GAGCATG AGCTGACCA ACCAAGCTA TGAAGAGAA TGAATGACTT TTAAAAATAT 120  
ATTACATTA ACTGTCTAG GATACTCTC TTGAGGCTT GGAATCTTC TTCTTGAAA 180  
50 TTTCATATC CACTCTAGT CTGTACAAA AGATTTTAT CTTGAGATCG CAATTTCTC 240  
TCTCCAGAA AAGATGACTA CAGAGCTC AAGGATATG CTTTGTGCT CAGGGATTA 300  
55 CACTATGCTT TTCTCTCTT TCAATGCT ATTTACAGA GACCTTCTA TCAGAGAGG 360  
TACTGACTA TCTTTATGAC TTGTGATTTG ATCAGAGCTT TAAAAAAA AAA 414

## (2) INFORMATION FOR SEQ ID NO: 18:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 469 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 18:

5 ATACCACTT GCATACAAA TGAATGCTT GAGCTGTACC CCACTTCTT 60  
CAGGAACTC GAGATGATG TGTGTTCTC CTTGTGTGAC ATAGATGGA TATATACCC 120  
CTTCATAGT TTAGGCTGA GGAAGTCTG GAAAGCTTG CCCCCTATT TTTTGGGCT 180  
GTTGGGCTG CTTCGCGGG GAGTGAAGCT ACTTCTTCA GAGACCAAG GGTTCGCTT 240  
GCCAGAGAC ATGAGGAGC CCGAGAACT TGGGAGAAA GCAAGGCCA AGAAGAACAC 300  
GATTACTCT AGGTTCAAA CTTGAGAAC CTGCGGACG TGAAGAGAT GTTTTGGGC 360  
GATGTCTGT TGAAGGATG AAGATGAGT TATCTCTGC AGAATTTCT AGAGCTTTC 420  
25 ACTTCTCTT ATTCTTCTC ATACTTCTT ACCCCCAAT TATATGAG 469

30 (2) INFORMATION FOR SEQ ID NO: 19:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 550 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 19:

40 CCCCCCCCC CCCCAGCT TTAGAGCTC ACCCCGAGC ATTTCGGGT GGTTCGGCC 60  
TACTCAGCC TGAAGTCCC TGAAGAGCC TCACTTCTT GTTCCCAAT CCCCCTACT 120  
GTCCAGGAT GCAAGCTGA GCTGGGTCC TTGTTCTGG CTTGCTGCT GCGCCAGCC 180  
AGAGCCGAG CTTCTCTGA ATTGATATCA GTCTTCTCT GAACTGCTC CCCCACGCT 240  
GGCATTATC CCAGAAACT TTATGTTTC TGAAGCTTA GCACTGCTG GACTCAGGG 300  
50 ACTGTGTGAG GTAGCTGAG TGGAGCTCA GTCTTAGAG GTCTCTGAG ATCTGAGCT 360  
AGGACTTCC TACTCCCAA GCGAGGCC ATCAGCGAG CTTCTCTGA CCACTGCCC 420  
TGTGAGTGC TGAAGTCAAC CAAGGCTG CAGCTCTG GCTTCATTA AGGATTCGT 480  
55 ATGAGCGAT GGGCTCTGA GCGAGCCCT TAAAGATCT GGTCTTTTT TGAAGAAAA 540  
AAAAAAAAA 550

60

## (2) INFORMATION FOR SEQ ID NO: 20:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 741 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 20:

5 TCTTGAAGG TGTACAGTAC AGGATTATTA TATGAAAT TTTATATAC AGGTTTCTT 60  
TGGCTCTCA TATATATTA GCAAAAGAA TTGTAAAT GGCACATAT TCGATTAAC 120  
TTTTCAGTG CGGCTTTCT TCTCTTCTT TAAATTCAAA CTTAGATACA TCGATTAATA 180  
ACTAGAGAA TCACTTTTAC CTTTGGGAC AGCCAGTTT TTTTATAAA CTTATTTCTT 240  
20 AGCATGCTT CAGCAAGTG TCCAGAGCC TAAATTTGTA AGGAGCCAT GTTCTTCTT 300  
TGTACAGCT CCTGAGCA TTGTTGAGG GACCAATGC ACATGCTTC ATGGGCAATG 360  
25 NCATGAGG CACTGGGTG ATATCTCTT ACATATTTG ACATGCTTC CTTCTCTGG AGCTATPAC 420  
ACAGGCTC TCTTTCAGT GATCATTTG AAAGTCCC CAGGCCCCAC CATCCATGT 480  
GAGAGAAA CAGAGACTC CTGAGAGA GTCCAGCTA CAGATACAA CCGTTGAT 540  
TGGCTGTC ACCTTCTCC TCCACTCTT GTATCTCTAG AGATGCTGG TGAATTTTC 600  
CTTAAGCTA CTTGACTTC CTGTAAATCT CTAACTGAG TTGAGGGCT CAGAGATG 660  
35 ATTGTAGAG TGTAACTCC CAGTAGGCT TCTTTATCA TTTGTGTGC TTTTCTCTC 720  
ATTAAAGAA ATGATTTTCC C 741

## (2) INFORMATION FOR SEQ ID NO: 21:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 991 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 21:

5 GGCAGAGTC TCCCTGGGG AATTTTTCT TTTTCAGAG GAGGAGGGC TTTCCAGGT 60  
AATGTTCTA GAGTTTGGG CAGAAATCT GGCAGAGC CAGACAGT CTTCTTTAA 120  
TCCAGTCA TTTGCTTCA TCCAGCTAT GTTTCAGTG TCCCTGGGT GTTTCAGCA 180  
GCAAGAAA ATGAATTAAT CTTGTGAG TTTTATTTT GTTCTCACT TGTTTTACA 240  
60 CTTATTTTC TGAATTTAT GGTGCTGAG AATTAAGAG GAAAGTAGA AATAGTTAA 300

ACTGAGTTG AAGCAATAT ACTAATATA GATAAGCTG ACCTTGAT ATAGCACTT 360  
ATAAGCTTA GAGTGTGCTA AGTTGAGTGC AATTTTCTT CTGATCTTTC TGATGCCAA 420  
CAAAAAGCA CTCATGTTTG TTATGTGATT GGAATGCAAC CCGAGAGAG AGCTGTCTT 480  
GTCTTTGCG GAGCAGAAAG CTTCGTGCA CCAAGTCTGA ACCACCACTT TCATGTGTAC 540  
ATAGATATG TGCTGAGCA TATTTCACAC CCGCTTGCA GTAAACACTT GTATGTGTT 600  
CGAGTGAAA CGGTCACTT CGCTAAGC AGCGGTGTT GTGAGCGCA AATGTCTATC 660  
TGCTGTAAA ACHAGCTTC CATGTATG TGATCTCTT ACTCAAGAG TGATGTGCA 720  
AACHGCTTT GCGAGTCTT CTGTATTTCA TGGATGAAC CTGGAACATC TTGAGACTG 780  
AGTTAAGAT AGTCTCTTAA ATACTCTCC ACAGTTTCTT CTTAGTTTAT CTCTACATC 840  
AGGCTGTGCA GCGGCTGTT CAAGTCTATA TTTTCTGGA AATATTTTCA GTGTTTATTT 900  
CGACTTATG CCACTCTG TGAGCTTAT TCTTCTAAC TCACATTTA TGTGATATTT 960  
ATGTAAATTT AGGGGACTG TATTGCTTT A 991

(2) INFORMATION FOR SEQ ID NO: 22:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 653 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 22:

CCACCGTCC GGAATTCCTC TGAGATCTT GGGCTATCTT TCACAGCGA TTCTTGCA 60  
TTGATGCTTT CTACAAATCA ATATAGTAC TCCCAAGA TGGAGACTT GATTTCTAC 120  
AGATTGATG AAGTGTGGA AACHACACAG ATTGAGATA CGGAACCTAT GTCTCCAGTT 180  
CTCAATTTTA AATTTGTTCC TCTGAATAT GATATATCC TGATGATCC AGCAGAGAT 240  
GTTGAGTAC AACTGATCA GATGATGAC AACHAAGG GAGTGTATAC AGACACGCG 300  
GATGATTTA GTATTTTAC CTTGTTTCC AAGCGCAG AGAAGAGGT AGCAGAGAA 360  
GTTTGTATG ATCTGCTTG TGATTCGGG AGTCAGGAG TTCTGTACCC AGCTACTGA 420  
TCTGAGCAC TTTCTATGTT GTTAGTAC GAGCAAGCTA TGGAAATTTA AGACACCAT 480  
CCAGAGAGG GCTTCTGAG GTTGAGGTG GAGAAATCC CTGAGACACC TTGTGAAGT 540  
CAAGGAGAG AACTCAAGA AGAAATATG GAGAGTTTC GTTTCACCT TTCTCTGACT 600  
GAACTCATGT CCGAGAGGTT GTGTCTCG AGGCATCCA AAAAAAAAAA AAA 653

## (2) INFORMATION FOR SEQ ID NO: 23:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1486 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 23:

GGCAGCTGA CCACTGCA GCACTGCA CCACTGCTG CTGCTCTGCA GTTGGGGAC 60  
CTTGGGAGG AAGCTGCTG AGCTCCAGA CCGCGGGGC CATGGCGCG GATCTGCTG 120  
TTGGCTTGG GGGCTTGAG CCGCGAAGC GCTTCTGCA CCAAGAGAG TCTCTTGGC 180  
GCTGGCACT GTTCTGCGA SGAATGCGA TTGCACTCAT GTTCTGCTAT GCACAGATCC 240  
TGTGTTTGG GGGGTGCTG GCTGTCTAT CACTGCGCA CTTTTCAGAC ACATTTTGG 300  
TGATTCCTAT CAGATTCCTG ACCATGCGT ATGCTGACCT GTTCCCGGC ACATGTTGG 360  
GCAGATCTT TTGCTGTGCT ACTGAGTCA TGGGTGTCTG CTGCAAGCC CTGCTGTTG 420  
CGGTGTTGC CCGAAGCTG GATTTTAA AAGCAGAGAA GCAGTGCAC AACTTCACTA 480  
TGGATATCCA GTATACAAA GAGTGAAGG AATGCGCTC CCAATGCTTA CAGAAAGCTT 540  
GATGTTCTA CAACATACT CCAAGAGAG AGTCTCATC TCCCGCAG CATCAAGCCA 600  
AATCTGTGC CCGATCAC GGTTCGCG AGTGTGCGCT GAACACCG AGCTTCGCG 660  
AACHGTGAA CTCTATGTT GACATCTCA AGATGCAT GATCTGTAT GACTGTGAC 720  
AGATCTGAG CAGCTCAC CCGGCTCTG AGAAGCAT TGACAGCTG CGGGGAGGC 780  
TGATGCTCT GACTGAGCTG CTTAGACTG CCTGCGGCC GAGGAGCTT CCAAGACCA 840  
GCCAGCATC CAGTACCTG GACCCACAG GAGGAGCAG GCTACTTTCC CCAATACTCA 900  
GTTGTTGAC ATGCTCTCT CCACTCTCA CCGAGCGCTG ACAAAGCAC CTCAGTCTCA 960  
AGAACCAAG GGGGCGCTG CTTGAGTGG GTTGTCTTCC TGATGCTGCT TGGAGGGAC 1020  
GCTGGTAAA GTGGAGGCG TTGCGCCAC CTGAGCGCCC AGTGGGAA ATGCTCACCC 1080  
CCACTCTCA TACCTCTAT AAAAACTC TCATATGCT GCTATGAGC ACCTTCAGCT 1140  
CTCAATTACA AGTCAGGCG ACTGAGCCA GCACTCTGCT GTCCCTGGA AAGAGGATAC 1200  
TAGGGGCGCG GATTCAGAT TTCTGAGGC TTCACTTACC GCTGCGCGAG CTGAAGACT 1260  
GGATATAGG CTGGGCGCG GCTGAGCTG GCGCGCGCTG GTGGACAC AAGAGGACA 1320  
CAATTTTTC AGAGTGTGAG AGAGACTG GTGGGAGGA AGAATGTATA CTCACAGCC 1380  
TCTCTCTTA TCTTTTAT AATGTTAA GCGAGAAA AATAAAAAA AAAAAAAAA 1440

AACTCGAGGG GGGCCGCGAC CCAATGCC TATAGTAAAC GTANN

1486

5

(2) INFORMATION FOR SEQ ID NO: 24:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2323 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 24:

CTTCCGCGTT TCTCTGCA GGGAGGTCC CGCTTCCCG TGGAGCTCC GACCAAGCC 60  
CCTTCAGCTT CTGCTCCCG ATCGATGCG TCGCGCGCC GCGCGCGCG TCGCGCTCC 120  
TTCCGCTCT GCTCCCGGA CCGCGTCCG CCGAGCCAG CAGCATGCG GGGATCAGA 180  
AGCAAAAGAC GGAGAGCAG CAGAAATCCA CCAATGTAGT CTATCAGGCC CACCATGGA 240  
GGAGAAATTA GAGAGGCAA GTGATTCGAA CAGCGGTGG GTTCCGAGAA TGTACCTGT 300  
GGCTAACAGG TCTCTGCT GCTGGGAAA ACAGAGTAA GTTTGCGCT GAGAGGTAC 360  
TTTCTTCCCA TGCATCCCT GTTAATCTT GGAATGGAG AATGTCCTTC ATGGCTTTA 420  
CAGATTCGCC CAGATGCTT CATGCGCCC AAGCATGGA AGTCTCTGAC AGATTTATAC 480  
AGTCTCTGCG AGAGACTA AGCTTTGCT CCAGAGTTC TTCTGAAAT GCTCTTTGAT 540  
TACCTTTTCT ATTATTAGA TTAGATGCT TGTATTAAAT TGTCTTCAA TGTATCTT 600  
TATCTTTTTA TATAGATA AAGTTGTGT CTATATTTAA AAAATATAT ATATATATAC 660  
ACACACAT ATACATACA AGTCAACTG AAGACCAAT CTAGCAGGT AAAGCATA 720  
TTCTTATACA TTCTATATA AATTAGCTC TATGTATTT CTACTGACC TGAAGAGCA 780  
GCTCCAGAT TTCTTAGGC TTGTGTTGAC CATGTGCTA GTTACTGCT GAAAGTGAA 840  
TATATTTTCC AGCATGCTT GACAACTGT ACTTTCOA TGTATTTAT CAATGTAAA 900  
ATATATACA TGTGCTCT TCTGTACAT TGACAAAA AAAATTTTT TTTTCTACT 960  
CTAAAGAGG TGTGGCTAC ATCAGATTC TTCTGATAT TTACTCTAT GCTGTACAA 1020  
GCTTAAATGT TATATATA TCTTACGT TGAAGCTG ACTGAGAA CAAATGTGC 1080  
AATAGCTGA ATTATATCT AGAGATCTGT GCAGCTATT TCTGTACA AAGTTATAT 1140  
GTCTATTAAG AGAGTCTTA ATGCTCTG TGAATATCT ACTTCAGTT ACAGGTGAC 1200  
TTTTATTAAG ATACAGTAT TTGATGAAG GAGCTCAAC AATGTGCAA TGTCTGGA 1260  
AGTTATCTTT CCGCTCTTT CTGTGTGCA TTGTCTTGG CAGAAAGAT GCGCTGTG 1320

CAGCAGCC GCGCTCTA ATAAAAATA ATTCACTA TCACTAGCC AAGCACTAG 1380  
AACTGAAA GACGAGAA AACAAAGAT CCAACCTTT CATCTTACAG GTGAAACAC 1440  
TGTGATGAG CACATGATG TTTTGTGTA GCTGTGACA CCGTACAAA ATTATATTT 1500  
GCAATTTTA GGAAGTCTG GTGGAGTGA AGAAGACCC AAGCCACTG ACTCCAGTC 1560  
TGTGCGCTG TCTACACAG AACACAGG AGCTGGTCA GATTCGCC AGCTGCTTAA 1620  
CAAAATCTT CGAACAAA GTCTTACAA AGCTGCCCTC TGGATACTG AAGGTGCGAG 1680  
TTTTCTGAC TGCATGAT TTATTCAGT TGAAMAAA AAAAGCTAT TCGAAAGAT 1740  
TGAAGCTTT CTGAGATTC TTCTGATGC TTACTTCT GAGAGCCAT GTTTTACTT 1800  
TATGCTAAT TCAATTTTC CAGGATTA ATGAGAAA CAGACCTTT TATATATAG 1860  
GTCTCTGCG AAGAGACTA AATTAGAAAG ACAAACCTT GCAATTTTC ATATCTCAT 1920  
TCTTAAAA CATTATCTT ACTTACAAA AGTCTTTTG AGATTAAGT ACACAGATG 1980  
GCGAGAGG TTGTCTTTA ATAGTATAT GCTTACTG ATGTATCCG TTACTGACTA 2040  
CTGCTTTTA AAAAANAAC CAGATTTT ATTGAAACA TGAAGAGGA TTATAGTCC 2100  
TTAGCAATA TATTTTCTA CTTAAAAAT ACATTTAAA GCTCTCTCT GCTCTAGTAC 2160  
CATGCTAGT GCATATGAT ATTCTATGT ACAGCTGATG CTGTCTCTTA TTTTATATA 2220  
TTTATCAGG TGAAMAAA AAAAANAAC AAAAANAAC AAAAANAAC AAAAANAAC 2280  
AAAAANAAC AAAAANAAC AAAAANAAC AAAAANAAC AAAAANAAC AAAAANAAC 2323

(2) INFORMATION FOR SEQ ID NO: 25:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 683 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 25:

GCGCAGGCC TGTGTGTCA TTTTCTGCT GTCACATGAC CTGACATGAG CAGCAGGCC 60  
TCAATGCTG ACAGCATTC CCGAGCTCG CAGCTGTGT TGTGTGTGAA AGAGAGAGG 120  
GGCCGAGAG CCGCTTTTG AAATGTTTC CTGTCTGAC TGTGAGACA CTTGAGAGT 180  
ATTGTGCTT AATTTCGAC CTGCTCTGT TTCTGTACA TCTTGGAGG GAGCTATGCT 240  
CAGACATGC GCGTGTCTA GAATGAAA AGTCCCGGT CTGTCTCTCT CACTCTGCT 300  
CTCATGGGG AGGAAAGAA TGGCTTTGT GCTTTTGTTC ACAGAGTGA TGGTGTGCT 360  
GAAGTGTGC ACATGAGCC TGTGTGAGG ACTGTGACA CGTTCTGAC TTGTCTCAT 420

5 CAGGCTTTC TGGTCTCAT AGGCTGAGC AAGAGTGGAA AGGAGGAA AGAGGCTTTT 480  
CTGACAGCA TTATATATA TATAGATTC ATTCAATCT GGTGCTCTG GGCACCTTC 540  
CCTGTGCTC AGTGACATG AGATGACTGA CTGCAATAC TTGTACAT TTCTGTGAG 600  
CAGCTACTA GGGGAAACA GATGTATGTC TATTCGGAT AACAGTAG ATTTCGCA 660  
10 CTAAAAAA AAAAAAAAA AAA 683

15 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2036 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 26:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

25 CTGGAAGG AAGCAATGC GATCTGCTC AAAAAACAT AATATCAATA AGACTCATG 60  
TATTCGAAA ACGATGTG AAGCAATCA CATTCGATG ATCTTTTAA AAGGAGGAG 120  
ACCTTAGT ATCTCAAT CTGAATTTCT ATTATTTCT TCACTGATA TAGAACAT 180  
30 GGTATCTGA TTATATAGA TATATTTTG GATATTTAC TATTTACTT CATTATGCTG 240  
GTAACTATG TAAATCTGT TATATATAC AACTATATC TTTTTTAAA GAGGAAGC 300  
35 TTATTTTTC TTGCAATGT ATGATTTAT CTACTTAGT GTTTTCTCT ATACTGTAT 360  
TAAATTTTTC TTATATGTA GTTTTTCATA AATTTTAAA CCGTGTCCC ACCTTTTTC 420  
GATCTCTGTC TTGCTATAG GTATATAGC CATTTTAAA ATCTTAAGC AAAAAATTT 480  
40 TATTTATTA AAGATCAAC TTTGCAATC ATGAGCTGT GAGTCAAT ATTTAGTAT 540  
AAAAAGCA GTCTCTTTT TTATATTTAC CATTGACC CCGCAATG CACTGTATT 600  
45 ATATTAAGA AATATTAACA ATTTTAAAT CTCAGATTA AATCTATTC ACTACATCT 660  
TTTCCCTCTT TCTCTGAT TTAGCATGT GTACTTGCA TCTTCAAT GTCTTAAGCA 720  
50 CAGTGTGTTT CTCAATAT ATCATATATG ATTTTATAT GTCTCTTTT ATCTATGTC 780  
GCTTCTTACC AAGAACATA GGAAGAAC CATGACTGT GTAGAGACA TGAACATTC 840  
CTCTCATAT GTTTTATTTT CACTGCTTT TGAATTTCA CTTTTTAA CAGAGCCAGC 900  
55 AAGCAATTA GATGTGCTG GGTGCTCTG TCGGCGGCG TTTTGCACC GAGCTCTAA 960  
ATCTGTGTA TTGAGGATC GTTTTGTGA CTCAGATTC GAGCTACAG TGGGCCCCC 1020  
60 TCTCTCCAT TCTTTTGA AGCACTGAG GGAAGAGG GTTCTTTTC AGGTGTCTTT 1080

GGCTGCTTT TACGGATGC GAGCTTCTC GGAATCTTT GTTCTCTGC ACCTCTGTA 1160  
GCTACTGCG GTGCAAGTT GTAGATTTA TTCCGAGA GCTGCGCTK GGGGCTGAG 1200  
5 CTGGGCTAAA TCGAAAGCA TCGAACAGA AGCGGGGCAA GGGAGGAAA AGAGGCGCTG 1260  
GCTCATGCG TCCCTGGAG ATGTCTGAG CATCACTGC CAGCTTGGC CTGGGGAAC 1320  
10 ACCGTAGCA AGGCGCTCAG GTTGCTCTT TACAGAGA ACCTCAGAA GATTAATTTG 1380  
CACATGAGC TGTGATACA CTATGTTCA TTTTTTTTT TTTTACAGT CATCAGGAT 1440  
GTTTCGAAG TGAATTTAT TTTTGTAA TTCTTTATC TTTACTTAA GGTGAATGTG 1500  
15 TATTCCTCG GAGGATAG GAGGAACA GGAATTTAA TATGTCGA CAGAAACTT 1560  
CCTGCTTAT TATATATA TCTCATGTA TTTATGCTT AATGTAGCT GACTTTTAA 1620  
AAGCTTCTT TTGTGCATG CCGTGTGAG GATCTGTAT TGTACATCA TCCCTTTCT 1680  
20 CCTGTTTCC TGTAAAGT TATGACAA AGAATATTT TTGCTTACT TCAATGTCC 1740  
AAGCAATCA TATTTTAAA ATTGTGATA TATGGAAGA GCAATTTGT TACATGTAA 1800  
25 AGCTTACTG ATATACAGT ATACTAATG TTGAAGATC TGTCTTTTC AAGTGTACG 1860  
TTTTCAAATG TTGTACAG TGAACACC TTGTGTTTA AACTCTTAC AATGTATTA 1920  
TATTCATTT CCGCAATGT AACTAAGAT CATGCTATA TTTCAATCA AGCTATATT 1980  
30 GAAAGTAG GGAATGAT ATACAGGT TTTTAAACA AAAAAAAA AAAAA 2036

## (2) INFORMATION FOR SEQ ID NO: 27:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

GGCAAGCAT AACTNAGCA CATTATACT GTATGCTAC TTCTAGATT ATAGAGATT 60  
AAGTTGAGA TCAATTTCC ATTGAGAG AATATGTTG GTTTCAGTC GTTTTATTT 120  
180 ATCTGTGAG AAGCAGACT CCGACCAACA TTCAACCCA GCGCTGATAT GACAGTATC  
CTGAGAGCA GAGCCAGCA CAAACAGCA ATGTAGAAA GTTACAAATG GAAGTTTTC 240  
TCCCACTTC GGAATGACA CTCGAAGCT GATGCCAGAA ACTGCCAGG TATTTCTCT 300  
55 CATTACTCT CTAGCCACC ACTTTAGCT CCGCAATTA ACTAGTCCAG TTGACTATC 360  
CTCTTACTT TTATCATTTA GGTGAGCAT TGCAGAAA CTCTGACTT TGCATATA 420  
60 GGGCTGTGTT TCTCTGTGTT CCGGATNAG AGGCATACC ATATCTCGA AACATCGAT 480

AAATGAGAT TCTTATCTT CTTCCAGAG CTTCTGAGTT AGAATTCAC AGTTCTCCA 540  
GGTATCTCA TACATCTTA AGTTTGAAA CCATGAGTA AGTTAATGC ATTAGAGCA 600  
GATTAGATG GAATGGTGC GTATCTTCTT ACATTTTCC TTTTACGAG AATGAGAG 660  
GTCACTTAT CAGACATAG ATTATTTATT GCTAAAGCTA AAAAAATTA AAAAAA 717

(2) INFORMATION FOR SEQ ID NO: 28:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 495 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

GAATTCGCA CGAGCAGAT CTAATTTTA GTTTGGAGT GCATCTTAA GATCTCTC 60  
TATTCCTTT TCTCCACA TTATCTCA TTCTGCTCT CTGTGAGT TTGCTAGG 120  
AATGACTG TTTTTCAT AGTAATGA GAGCTGAT TTTTTCATTT CTGAGAGA 180  
AGGCATTT CTCTACAG CTGAAGCA CCGCTGGTG GCTGGGGCC TGTGGAGT 240  
TTCTGGGGA TTGACCTTA CAATGCG TGCCCTTCA GAAAACTG CAATAAAA 300  
TTATTTTAA AAGGCTCC TCCAGGAT GCATTAAG GCTATCACC CATATTTT 360  
AGCTTCGAA GAGTAAAT AMCTCGAG AGAGAACTG AGCATGTA GAGGTTGGA 420  
ATGACTCAT GTTGACAC TATGGAGT TCCCCAGAG TAGCAGAT CAGCCACTA 480  
CACAGAACT CTAG 495

## (2) INFORMATION FOR SEQ ID NO: 29:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 556 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

AGCTTACGT CATGATCAT TAGGGGATG CAGGGGAAA CGATGATG AATGCCCCA 60  
GACACCTCTT AGAAGCTG CTAGAAAGC AGACAGACC AAGCCCTTA ATGAGATGG 120  
GGCATCTGTG CTTCTCTCT GCTACTGCT AGGGGTGAG CAGAGTGGTT CATCTGGGA 180  
CAGTTAGCTG GACATACGT GGACCAACA CAGCATTTT CTGGTACTT TACCAAGAG 240

ATAGAAAGC AGGCAGTCT TTACAGAGC TCTTACTCT TTGAAACA ATGAAATGC 300  
CCATCTTCC AAAAAAGT ATGTGTCTG CTTGTGCTT GAGCAGAT GTGGCTGAG 360  
GTCAAGATC CCACTCTCA AAGAGGAGC CAGTACAGC GCTGCACT GTGTGATTC 420  
ACATCTGTA CATTCTGAC AGGCATGCT TCGATGCAA AAGAGATC GGGCTGAGG 480  
GACTCTGAG AAGGGAGAG GGGCTGCTG GATGTGGTT GATTTAGCA GTAGCTCAT 540  
GAGATGTGAC CTCAA 556

(2) INFORMATION FOR SEQ ID NO: 30:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 434 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

CTAATGTGT ACTGTGCTT TTGCAACA GGGCCCAAT GTTAGGTGTG AACACAGAT 60  
GCACAGATG AGCAGCATG CAGATGCTG AATTAATCT CTGACAGAT GTGTACATG 120  
ACTTCTTTT CAGCTTATT CTGTGGCTG CCTTTGAGA TAGAGCTTTG TTGATATTA 180  
CATTAAACA AATTGTAAA TTATTTTCA TTCTGATG TATTATACA AAGGAAAT 240  
GATTAATCT ACATGAGAT CTTTAGTCA TCTTAAGAA TTAAAAAT CTTTGGGGA 300  
ACATTTTTG TATACATAA TGTTTAGTA GAATATTTA TAGATCTCT TATGTGATTA 360  
TTAATCTCC TATGTATTT TATATCTGA TGTGTCAAT TTTTATTTA TATGAAATC 420  
TATGAATAG GACA 434

## (2) INFORMATION FOR SEQ ID NO: 31:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 715 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

CCAGGCTCC GATCTCAG CTCCAGACT ATTCCAGCC ATACAGAC TGGTGTGAG 60  
AAGCTTACT CCAACTTAG CCAAGATGC AAGTTTGT TCAATGGGG TTAGACAGCT 120  
ATGACTATCT CCAAAATCA CCGCTGGAT TTTTCCGAG ACTTGTGTTT ATTGTGTTT 180



CTGGCTTAT TGGACTCTT TGGCTAGAG GTTCAAAAT AAGAGACTA GTGTATCCG 240  
CTGGTTTAT GGAATTAAT GCTGCTCTT ATATCCACA AAGGCCATC GTGTTGCC 300  
AGTCACTGG GAGAGTTA TATGCTGG GTTACGAGG ATATATAGT ATAGAGATT 360  
TGTGAGGA GAATTTCAA AAGCAGAA ATGTGAGAA TTCACTGGA ACTATAGAA 420  
AACTCCAG CTCTGCAT TTATCTAGT ATAGTTAAC ATGGAACTC CATAGATTA 480  
ATAGTATT CTACAGAAA ATGCAATAGA ATCTAGTAT GAATGATTA ATGTGCTTT 540  
CTTCTTAG AAAATATAGA CCGAACTCT GTTATCTCT GTGAATCAT CCTACAGCA 600  
AACTAACCTG GAATCCCTTC ACTTAGAT ATGTACAGG CCTTAGACT CCTCATCTC 660  
ATGTTCTAT TTATGACTT ATTTAAACC CAGTTAAA AAAAAAAAAA 715

## (2) INFORMATION FOR SEQ ID NO: 32:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 486 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

GAGCAGTGC CGCGAAGG GAACTTCT CTACTCTCG CCACAGACC TGTCCGAGA 60  
CACTTCTGC CCGTCTCTG CTGGAGGCC ACTTCTCTCC CCACTCTCG ATTCACCCC 120  
CAGTCACCC TCAACATCG CCGCTCTCT CTCTGCTTT GCGCTCTCT GCTCCCTGA 180  
GCTGTCTG TCTTCCCTC TTGAAGCA ATCCAGCTT CCTGGATCT TCTGCAACT 240  
CCACTACCA TCGCTTTC TCTGTCAG TCACTCTC TCACTCTC AAGGATATG TCTATCTCG 300  
GTGCTCTCT TCGTCTCTC AACTCTCTA CCGTCTCTA AGCTGCTAT TCGCTCTCA 360  
CTGCAGAAA CGGTGCCCC ACCACTGCG TTACAGGGA GGAAGCAGA ACATGAGAA 420  
AAGAAATAT AGGGCTACA AAGATCTCA GCTCTGATC CGAAGGAAA AAGATCTTT 480  
GGGCAC 486

## (2) INFORMATION FOR SEQ ID NO: 33:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 725 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

GTTCCTCTGG TATATATAG GTTATTCOA GAGGCAAT GTCACTTTT AATTAAGAC 60  
TTTCTGTTT AAGCTTTTC AAGAGCAG ACCACTTGA AGATTCGCC TAGGCTTGT 120  
ATGTGCTAA TTCAATTTT AAAATATAT CTGTCTTCA TTTTAAGCT TTGCTATAT 180  
AGTCAGAAAT GTCTAAATA ACAAACTAT TTATTTTAA TTTAGGAG ACTAAGGGA 240  
AAGAAATGA AACTCAGTC TTATGTMAG CTCAGAGAT ATTAGGCTT AAGGCTCTT 300  
TCTAGTTTAA TGAATATG TACTACTGAT TTTATATAT TCTGTTTAT GATGATAGA 360  
TCTCTGGGA AATGTTGAG TTACATGCG ATTTCACTGT GATCCCTCTC AAGCTAGAT 420  
CAATCTATA ACCAATGAC AACCTGCTC TTGCTTTAC TGTCTGTGA ATGTGAGCT 480  
CAGTTTCCC AAGATCTCG TGTATGAT GAGTCAGAT GCTTTCTCTC GGTGAGAG 540  
TTCTGCCCC TCTTAATTT GGTGATGTC CTTCAGATA TCTAAGCTC CAGTCTGATC 600  
TGTATGCTT ATCTTACTG TTAAATGAT TATGATAT GTTCATATC TTGCTTAG 660  
GTTCATATT TTCAATTTA TGAATAAA GTTTTTTCT GCTTATAAA AAAAAAAAAA 720  
AAAA 725

## (2) INFORMATION FOR SEQ ID NO: 34:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 437 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

CACACAGAT GCTGCCCTCA GACTGTCCA TCTGTACCA CATGAAGG CTGCTGCTC 60  
TGCAGATAC TGAAGATTC AAGCTGCTC TCGAATGTT CCAAGACAT TGCAGATC 120  
CTCTGCTTT TATGCTCT TGTGAAATC AATGTGAG ATCTTCACT CCAGCTGAC 180  
CCAGCTGGA AAGATATCC AGGTCCATCC CGAAGGACC AACACCGAT ACATGACTC 240  
AGGAATCTTA TAACTTAAT GACTCTTTC CATCTGTACA TTGTGTGCA CATGCACTC 300  
ATCACTGCG GTGCCAGAT CTTGCCAGG CAGACCTG TGTATATCC AGGTATCTT 360  
CTACATCTG AGCTTAGGT TAGCTGATA TCAATTACA TTCTACTAT AATCAAAAA 420  
AAAAAAAAA AACTCA 437

## (2) INFORMATION FOR SEQ ID NO: 35:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 943 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

5  
GGACGAGCT GGACGAGCA CTAAATCCCA CGAACTGAC ATTGTTAAAC ACCTAAAC 60  
AGAGTACTT ACCTCTGGA GATTATAT ATATGGTGT ACATGATCA TGTACATAT 120  
GATGACCAAG ATGCTTATGG TCTACATTTT CTTTATCTT GTTATATTA CTTCTCTTA 180  
TCTTGTCCA TTACATGCT ATTTCTCTT CATGTATTAT TTCTTAGTGA CAGATCTTA 240  
20 ACATTTCTTA CAGCTTGCCA GAAGGAGAG AATGTGTTT TGGGTGGGT AACTAAATTT 300  
TTGATGAAA TATCATAGA TGAATGGA AAAAGGAGA CAGAAATAT TATACAAA 360  
AACAATGCT TTTTATGCC ATTGACTGG CTCTTAAAT AGTCTACAG ACATTCAGT 420  
TTACATCAC TTTTATGGA ATAAATGTC CACTACTAGT ATGCTCTTA AAGGGGAAA 480  
TGTGCTTAG TGGCTAAGG CTAAATTTG GTCAATTTG ACAGAGATG TTGTATATAT 540  
30 TGCATTAAT AGCACTAT TTTCATPAG TGTATTTT TGTATACAT TTTTATACC 600  
ACTATTTGT TGTAGCTTT TTGTTCTTN AGTTGNAAN ATGACATGC TNAATTCAAA 660  
CAGATTACC AATTCAGAA CTAGGAGAG CAATTTATGT ATGAAAGAAA TTNATGAT 720  
35 NOTCAATTC AACCTTGA TTAAAGCTTA GACTAAATAG TAATATATG TGGATAGAT 780  
TTTGCTTTTG TGTATTTTAT GTGATTTAG GATATAGAT TTACCTTTAT TTTGTAGAA 840  
40 AGTATTTGT ATGTGTTTA TTATTAATTA AACTGTACC AGGAAAAAA AAAAAAAAN 900  
NAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAA 943

## (2) INFORMATION FOR SEQ ID NO: 36:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 604 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

5  
GGACGAGAA ATCTTCATGC TGTATCTACT CCAGACCATG GATGGCTTT CCAGCTGAT 60  
GATTCCTATG TCTGCGCTGC AGTGGTTGG TTTTCAATGT TCTTCTTAAT TTTTCTTCA 120

TTGATCTTG GAGTTTCTT TTGTTGCTC CTGTGTTTGC CCAGCTTTAA TAAACACAG 180  
CCAAACAAA AACCATAGA TTCTGAACA TAGGGGGGCC ACATGGAGC CATATATGTA 240  
5 CTTTATGGA CTTCAAGAAA AATCTGAT CGGAAAAATG ACCTAGGAA TGTATACTGC 300  
ACATATTTA TGGCATATA TGGTGTGTT TCTTAATTT GTTCTCTTG CCGAAATGTC 360  
CCTTCAAT TAAATGACC TTTTCTCTT TGAACTTTT TTTTCTGCT TGTATATTA 420  
10 AGGTTTGA AGATTCATA ATTCAGAG AGTTTTCAA CAGAGATA CAAAGATTC 480  
TCAGTATTA TCTTCTCAT GTGCTTTAC ACCCACTAC ATTATAGAT GTATATGTA 540  
15 CAGAAATAT ATGCTGTGT ATGTGTCTT ACTCAATAA GTACATGCTT CCGAAAAA 600  
AAAA 604

## (2) INFORMATION FOR SEQ ID NO: 37:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 349 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

GTGATGCTC GAGAGCCCG AGGCGCTGCC CTTAAGAGG ATATCTTTA CCGCTGCTT 60  
GTCCACACC TAAACCCCA GCTCTCAGG CATGTGGAC ATGCGAGGG CTTCACTGG 120  
GGCATTAGA GCATTTGGG GACTGCGAT GCTCACTTT GACTTCTGC AGTTCGGGG 180  
AATCCAGAT CATGATGAC AAGTTTACA TATCTGAT CTTCAATGT CTATCTCTC 240  
40 CCGCTCGG TCTACAGG TATATGCC CACTTCTGT TCTAAATTA GATTAAGAT 300  
CATCTAGA GAGCTGTC CCGAGAAA TCGAAACGG TTGCGACA 349

## (2) INFORMATION FOR SEQ ID NO: 38:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 672 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

GTATGCTTGG CCGTTGGCG GATGCGAG ATCTGCGCT TTGAGCTGTT AAAAGCCAC 60  
TGGTACCGG TCGTGTGCG TTGTGTGAT GTATGCTTG CTACAGAGT GATGTGTGCA 120

5 GGAAGCGAG CAGCGGTAC AGCGAGTAA ATAGTAATGC CGAGCGAGT TTCTCTCGGC 180  
TTTATCATGT CAGCAGCTGT GGTATATGCG TTGTGTCTCG CGAACTTTGC GTTGAACTAT 240  
TTTACGATTA ATCAGATGCG GCGTGGCGCA ATATTCAGTA TACGGCTCG CAGTGTGCG 300  
GCTGATGTT CAGTGGCTGC GCGACGTTT TTGCGGTATG TTGCAACCA GGNCTTTTAA 360  
ACAGTTTTC GACCGGTTT AGCTTCAGCG GTTCATGCG CCGCGGTAGC TGTCTCTTAG 420  
GTTCACGCG AGCATAGCA TTAACTATCT CATCATTTTG CTCTCTGCTG GCGCATATCA 480  
TACTTTTCAG CATATTTTAA CCGTGGCGCA AACGGTTTAG CGTTTGGCCC ACGGATCAT 540  
AGCGATGCG GTTATATGTA TATCATATTA CACACACAG TACGGCTTCA GACACGTTT 600  
CGATATGCT GCGTCACTG GTTAACACCA CCGTGGCGCG ATATGCGCG TGCACAGTT 660  
CATCGATTA AT 672

## (2) INFORMATION FOR SEQ ID NO: 39:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1908 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 39:

5 AGATGTGTA TTTTATGAA CAGTATTTT ACTTTTAGG AATTTGCTA GCTCTTGAC 60  
TTNAGAGCT TTAGGAGCTC AACTTTTCTT TGTAGAGAC GTTCTTTT TTGATTTTA 120  
CAGTATPAA AACTTTGCTT TTGTTGANT GTATAGGTT AANAAGGAA TACTGTATG 180  
CAGGTTTCAA AAGGAATGT GCTTTAGGCA TGAATCAAA GATCCATG TACTGTAGG 240  
CATTTTNTT TCTTTTNGA ATGACATCA GCTCTTCTCT TCTGACTGCT AACACATAGC 300  
CCCAAGCAT GAGTTATNTT TCTATGCGT TTTTATTTT GTTTAGTTT GTTTCTTAC 360  
GCCAGCCAG TCTCTCTGCG GAACTGAC TCTGCTCTCT ATGAGACA AAGTTAGAAA 420  
TCTCCGATA ACCTAAATA ATTAGAAAT GATTTAAAA TGTGAATCG GGTAAAGTG 480  
ATGATGATA ATTAGCTGC AAGAACAG CTTCTTCAT CAGACTGCG TACTGTTTC 540  
TTCTGTAGC ATTTGTTTG GAGAGCTC TTGTTTCTT CTCTTTGCGG TATGTCTCG 600  
TTTCTTATA TCTTTGTAC ATTTATGCA TATATTCAC ATACCTTCA ATTCATTTT 660  
TTTAGGCTA CAATTTATG GTTTTATG TATTCAGAA GTTGTATAC COTGACACA 720  
GTCAATTTA GAACTTTG TTACCCAAA AAGAAAGCT GTACCTTCA CCACTCACT 780  
CTCATTTCT CCGAGTGGC ACCCATGCC CGAGCCCCG GAACACTAA TCTATTTTC 840

5 TCTCTTACA TTCTTTAT CTGCTCATTT CATTAATG GATTTCTACA ATATTCGCTC 900  
TTTTCGACT GCGTTCCAA ATATGATTTT CTATATGAG TGAAGAAAT CTCTCATCT 960  
TGAAGACTT TATTCCTGTG AAGCGAGTG GTTGTAAAA TCATAGATTT TCAGCGAAGA 1020  
GGCGCAATA CTTACACAGT TTCTTCGCT GATCTTTATG CTGAGTAGTT TTTCCTATA 1080  
ACCAAGCAT TATCATATAT TACTACTTAT ATATCTGTG CTAGTCTCTA GATTCGATGT 1140  
TGAATCTTT GCGTCTCTAG TCGGAGAGG TCTCTCTAAA AATCAGCTA AANAATCAGC 1200  
CAAAAATCAG GCAATATCAC TTGCGAATA ATTCACAAAG TCGTTTTCAC GTTGTCTAT 1260  
CTTTCTAGC AGCTTTTATA CTTAGGCCA GGTAGGCTCC CCAATTTTCT TTTTTCATTT 1320  
ACTCAGTGA GTTTCTGCTG TCTTTTCAA GTATGACCA TAGGACTTAA AGTTCATTTG 1380  
GATCGTTGT AACACTGCTA ATATGCTAA GTACAGAAAT TTATCTACAG TACTGTGACA 1440  
CAGTCAATTA TTGCTTAGGG TAGTCAAAA ATATGATG AGCTAGTTAA GCGTTTCTTT 1500  
GACTGATTC AGTATATTC AGAATGTGT ACCAATCAG GCTCTTTTAA ATACGAGCG 1560  
ACTCACTAA TAAACAGGA ACCAGCGAA TACTGTGAG CCGCAGATA TCGATATCAA 1620  
TGAATTCAG GTATATATC TCTGTACTC CTAATGAT GTTTTCTTAG CATGTGCTT 1680  
TCTCAATGCG TTGACAGCAT GTTCTGCTT GTATGCGCC TTTTACATG ATGTATGAA 1740  
TATGTGTTT TGTGTGAGC ATCAATGCT GTAAACCAA ACTAAACAG TTTTCTTGG 1800  
ATATGTTTC ATCTTTAAA TGACCTGCC CTGTCAATA AATAAATGAT TGTCTACCC 1860  
TGTAAAAA AAAAAAATTT AAAAAACTG GAGGCGGCG CCGGTACN 1908

## (2) INFORMATION FOR SEQ ID NO: 40:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 458 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 40:

5 CTTCAAAA AAAAAAAGAA GAAAGAGT CTTACAGAA GCGGTGAT CTTCGTGCA 60  
AGGATPACA TGAAGAAAT TTCAATTTCK GCTATTAGTT TCCATTCCT TCCCATCCA 120  
GCGTAAAGA GAACAAAG AATATGCG TATCTGTGT GTCTCTAGCT TTGGCACTTT 180  
TGTGATGTT CTAAGAGAG AGTACCTTG CTAAAAAG TCATTAATCC ACCCACTAA 240  
TAGTAAGCT GGGAGGAAA TGAATTTTC CTTTGTGAT CTCCCAANT CCAATTTGT 300

CACACAGCCC TCCACAGACC TCTCAGTTC CTTCACAGTC CAACTCTGTC TACTTGGCCC 360  
GCACCCGAT ACTATTGTGC CTACATTCAC CTTCACATGG CAACTGGCCT CCACTTCGCA 420  
CATAAACCT CATATTTTA ATTAAGTTGA ATTTTGA 458

5

(2) INFORMATION FOR SEQ ID NO: 41:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1153 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 41:

GCACAGAGC CTCCGACCCA GGTGGTCTGG AGCTGGCCGG GAGATGGTG GATCTGAGA 60  
GGCTGGTCTT GCACTGTGGT TGGGGGAGGT GGGAGCTGTT TTAACTGTTT GCGCCCTCTC 120  
CTGTGCGGCC GTGGCGATCC CCGCGGCGAG TGGACCCCGG GCGTCTCTCC AGCTTCCGAG 180  
TCCAGCGAGC CTGGGGGGGG GGGCGGCGCC GAGACACCGG AGAGTTCCTT TCTTCCCTGG 240  
TTAGCTGGAC TGTGGAGCTG GTCTCTGTGG GCTCAGAGGC GTGGCGAGGT TGAAGCTTAC 300  
CTGGCGAGGT GCGACGAGGG GTGGAGAGGG AGGAGGAGGG GCHTGGAGCG AGCTTGGAGA 360  
ATCGGTGCTC CAAACTCTAC ACTCAAGGAT GCACTGGCCA ACTCTGGTGG CGATGGCTTG 420  
GGGCAATGTT CTTTGGATTT CTACACAGAG AGAGTGTCT GCTGGCCATT CTGACGACAG 480  
TGCATCCCAT GCGAGTCTTG GAGGCTCAGG GTGCATGTGG TAGCCCTTGGC CACGGAGCG 540  
GAGCGGCAGA TCTGCCGGGA GAGGTGGGT GAGAACTCTT GCGAGAGAT CATGACATTC 600  
GTGGAGGTGA TGAATCGCCA TGAATGACTTG CCGAGATGC CACACAGTTC GAGGTGGAT 660  
AACTGTGTTG AGACAGGCTT CCGGAGCTTG CAGCCCTACC TGTACAGAT CTCTTTCCAG 720  
ATCACTGATG CCGTGGGCAC CTCACTACAC ACCACATGC GCAAGCTCAT CAAAGACACC 780  
CTGCGCTCTG AGCTGGCTTG GATCTCTTGG AGCTCTTTGA TGGCTTCCAG AGCTTGGCTT 840  
TTGGGAATTG CACTTTTGGG GCTTTGGGCT CTGGAGCTTG CTCTGGGTCA TTGTGTGAGC 900  
TTGGAGAGGG CAGCGCCCGC TGGTCTCTTG GTTTTGGTGT TGGCAGCTTC AGGTGATCTT 960  
TTTAATCTTT GCTCAAGGTT CAGTCTGGCC TCTACTGTCT CTCCATAGCC CTGTGGGGT 1020  
CGCCCTCTCT TCTCACTGT ATGAGAGAGC CACACTTGG ATGGGGATA AATTTGAGA 1080  
CATGAGTTTG GCGTGAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1140  
AAAAAAAAA AAA 1153

60

(2) INFORMATION FOR SEQ ID NO: 42:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1983 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 42:

GGCAGAGAG GGGCGAGGCC GACAGATGT TCTTCTGCC TCTTCCGGT GCGGGGCCAG 60  
TATGTCTCG AGCTCTGCC GTAGAGCTT TCGGAGCGCG GATCTCTCC ACCGAGACA 120  
TGACGAGGG CTTTATTTTA GGAATCTATT CCAAGAAAA AGAGATGAT GTGCCAGAT 180  
TCCAGATGC AGGAGAGAT TTGATTAAT TGTTAGCTCG AAGCTGAGA GAGACTTTGA 240  
ACATATCTCG ACACCTCTCG AAGGCAAGCA AGACTCGAC CTTTATGCT GTGCATCAG 300  
ACTTCCGAG GGTGGTCTTA GTTGGCTCG GCAAAAAGGC AGCTGGAAT GACGAGAG 360  
AAAACTGCA TGAAGCGAAA GAAACATCA GAGCTGCTGT TCGAGCGGG TCGAGCGAGA 420  
TTGAGACTT GAGCTCTCG TCTGTGAGAG TGAATCCCTG TGGAGAGCT CAGGCTCTG 480  
CGAGGAGGC GTTCTCTGT CTCTATGAT AGATGAGCT AAGCGAAAA AGAGATGCG 540  
CTGTGTGCG AAGCTCTAT GGAATGGGG ATCAGAGAGC CTGGAGAAA GGAATCTCTT 600  
TTGCTTCTCG CGAGAACTTG GCAAGCCAT TGAATGAGAC GCGAGCCAT GAGATGAGC 660  
CAACCGATT TCCGAAATT ATTGAGAGA ATCTCAAAAG TGTATGAT TAAACGAGG 720  
TCAATATCAG ACCGAGTCT TGAATGAGG AAGAGCCAT GGAATCATTC CTGATGTGG 780  
CGAAAGATC TGAAGAGCCC CAGTCTCTCT TGAATATCA CTACAAAGGC AGCCGCAATG 840  
CAAAAGAAC ACCCTCTGTC TTGTGTGGA AAGGAATTAC CTTTGAAGT GGTGTGATCT 900  
CGATCAGGC TTCTGCAAT ATGAGCTCA TGAAGGCTGA CATGGAGGA GCTGCAACTA 960  
TATGCTGAG CATCTGTCT GCTGCAGGC TTAAATTTGCC CATTAATATT ATAGGTCTGG 1020  
CGCCCTCTTG TGAATATG CCGAGCGCA AGCGCAAGCA GCGCGGGAT GTTGTTAGAG 1080  
CGAAAGAG GAGAGCAATC CAGTGTGATA ACATGATGC TGAAGGAGG CTGATAGTGG 1140  
CTGATGCTT CTGTTAGCA CACAGTTTA ACCCGAGAT CATCTCAT GTCCGCACT 1200  
TACAGGTC CATGATGTA GCTTTGGGAT CAGGTGCTAC TGGGTCTTT ACATTTAT 1260  
CTGTGCTCG GACAAATC TTGAGGCGCA GATTTGAAC AGGGAGCT GTCTGAGGA 1320  
TGGCTCTCTT GGAACATTA ACAGACAGG TGTAGATGG CCACTTGT CATTTTACA 1380  
AGTTTGAAA ATGAGATCT CAGGAGCAT GTACAGCTGC AGCATCTCG AAGAAATTC 1440

TAACATCC TAACTGGCA CATTAGACA TACAGGCGT GATGACGAC AAGATGAG 1500  
TTCCCTACT AGGAGAGGC ATGACTGGG GGCACAGAG GACTCTCAT GAGTCTTAC 1550  
TTCTTTTAC TCAAGACAT GCTTAGTCA GATACTGAA AATGCTTCA CTCGTCTTA 1620  
AATTGGAG TTGACTTAA AGGTTTTT AATTAATGA TGAATCTT TTACGGAGA 1680  
CAAGGATG TATTAAAA TGTAGACAC ATGGAATTT GTATGCTTG ATTITTTTT 1740  
CAITTCAC AAGATTAAT AAGGTAAAG TTATATCTT ACTGTATAG GATTITNAG 1800  
ATACTCTAA AATGATAAA ATTITTAAG CTTCCTAATC ACTTTTACA GTATATGTT 1860  
TTCAITGGA ACAAATAT TAACTAGAT TTGTATGCT AGGAACATG GCAACTGAA 1920  
AATTACTAT GACTGTGAG AATGATAAA TGCACCTGT TGTCAAAA AAAAAAAA 1980  
AAA 1983

## (2) INFORMATION FOR SEQ ID NO: 43:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1406 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 43:

ATGATGACA CTTTGAGAC GATTATATC CTCCTCTCC ACCTAAGGC CTTAGGTTA 60  
ATAGTTGGA AAGACTTAT AAGATATGAC ATTCTTCAA GGAGAGAGA AGATCATCT 120  
TTAAGGCTTA ATGCTTATC NCTGTGCTT AACTGACCT GCGATTAATA CTTAAGAAA 180  
ATAAATAAT CAGTCAAT ATCTCTAAC TGAAGTTTA GTGGAGGAC TTCTATATGTC 240  
CCTTCACTTA TCACTTACT ATTTAGAAA GTGTACAGCA TACTGATCTA ATTCTTAAGT 300  
CTGATTTGTG CAATTTTAA TGTACTTTT TAAATAGCT TCTTACGTGC AATTCATAT 360  
TAGAGTAAA GCGCTGTCT AATTAAGG CTCAGCAAA ATTGTACAT GATACCACT 420  
TTCCAGAG GAGCTTAAA ACGTAAATGT GCTACAGAG TTITTTTAC TGTAGAGCA 480  
TCAGCTGCT CTTATATTA TGAATAACA ATATTTTAA ACAATATATA GTAGACAT 540  
ATTAGGTT TCTATATGC TAAATATAC ACGATATGC TTTCGCTTT TGAATTAAT 600  
GCTAGTGT TTCCCTTGG AATTITGCA GTTTCACAT GTTTCCTGG CCAGGTATAC 660  
TGTITGTGG CTTTATTA ATCCAAAC ATTGTGTGG ATCAGATG GTTCTTAAA 720  
AAAAAATA AAGACATAT AGCTAGAG TCACITTTCA GTTCATATA TGTACCGAG 780  
GTAGCATGT GTGGATGAG GTTCAATCA GNCATTTA TTGCTTGTCA TGTAAATTA 840

AAACCTTGA TTAACTCTT TTCAATCTT TTAGTAAA TTCTCTTT CAGAAATGAT 900  
TGTGCTTAT TTTTCAAAA ATTCTGTG AACAGCTCA TCAACAGAG CACATTTAT 960  
CTATGACT ACAGTATCT TAAITGTTT CTTCAGTTT TCTGTGAC TTGTAAATG 1020  
CTACAGGA TATTAAAA ATCTATTCAC TTAACTTAT AATGTTTAT GAATTAATA 1080  
CATGATCAC AGCTTTGTT CTGTGTAC CTATAAAA AGTTTCTT TGAATTTCA 1140  
TGTAAAGAC TGAACATAT GTATATGTT TAAATATTT TGTGTGTGA GAATTTTTG 1200  
TCAATGAAA TTAAAGAAC TTACAGGAA GGTTTTAAG TTAGAATAT TCAATGCAA 1260  
TAAATAGA AATTATTAAT ATATATTTT AAGCTGAT CAGTGGAT CTTCGTATG 1320  
TATATATTA GTTATATAT AACACAGAG ATTTTATG GTATTTACG TAAACAAAG 1380  
GGCGATTA NTGTTTAAA GNCCTTG 1406

## (2) INFORMATION FOR SEQ ID NO: 44:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1391 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 44:

GGGCTTAG GCGGCGCC AGTCCGAGC AGTCTGCT CTTCTGCG GCGCTCGGC 60  
CCCGGCGTC GCAATGACA GTGAGCTGA CATCTTCTG GGAACACA CCTTATGCA 120  
CGAGAGCTG TATCGCTCT GCTCGATGG TTACTGCTG ACCGACCG TCGCCTCTG 180  
GCTGCGCTG GGAATCTCG ACAGACTCG CCGACGCGA GCGTGTCTG AGAGCGAC 240  
CATGACAT TACCGACCT TCCATGCT CGAGCGCTG CTGATGCG GCGCCAGCT 300  
ACTGCAACG CTCATCTTC AATTTGCG CCTCCGCGG GCACTACTA TCGAGGTTA 360  
CTATGCTTT GATAGGCT TTGTTGCGA GGTCTGCGC AAGAGCTGT CCAAGGCGC 420  
CAGGAAGAC CTGATGACA TCAAGACA AAGAGGATC ACCCTGAGA GTCGCGGAG 480  
ACAGTTGAC AACTTAAC GGTCTTCAA GGTGTAGAG GAATCGGG GCTGCTGCT 540  
GGCATATTT CAGCAACT TCTCTCTCT TGCAGGTTG GCGAGGACT ATCAGACAT 600  
GCTCTCTTT GCTAACAC GCTTTGAGC AAGGAGAAA AACTGCAAT ATCTGACTT 660  
CGTGAATTT GCTTCTGG CTGAGCTAT GATCAAAA TGAACCTTG CACCCGTGA 720  
CTCAGATG GATGACATG ACATGACTT AAGACGAT TTCTCAGA CTTGAGAGG 780

CTCAGGTC TAGTGCTCA CAGAGACTT CTGAGCTGC ACAGAGGCT GGTGTGACT 840  
GCTCTCGGG AAGCTGGGC GTCTTCTG AGATGAGGC GAACTTGAG AACCTGTCC 900  
GGGGCTGCT GAACTGGCG CAGAGTAC CACATATA GATGTGAG ACCGTGTTT 960  
GAGCTCTG GAGAGTTT TCGAACCTG CCGCTCGAC CAGTGGCAC TCAAGGAGT 1020  
GGGTTCCTC CTGATCACT ATTGAGGTC TGTGATGCC CTGATGCT TCGAGACCA 1080  
GGCTCTCTG GAGCGTACA TGGGACCTT CCGGCTGTC CTCTGCGCC TGTATCATA 1140  
CTGAGTGGC TCCGACGTC GCGGACGCT GACATTAAG TTGCTCTGAG TTGTGAGCT 1200  
GGTCTCTGCT CCGGAGACA AGTGGGGGC GTGAGATGT GCTGTGTCT GTCTCTGAGC 1260  
ACCTGTGTC GTGTACAG GATGATGTC TGTGTGCT GCTTGGGAC TGAAGCATAT 1320  
CTCAGGAAT GGTGTCTG CTGAGCCAT CACCGGAG AGTGTGCTA CAAAAAAA 1380  
AAAAAAAA A 1391

## (2) INFORMATION FOR SEQ ID NO: 45:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1569 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 45:

GGCAGAGTG GAGATGCTG CCGGCTGCG GGGATGCTG CGAGGGGTC TCTTGGCCA 60  
GGCGGGCGG CTGCTACCC TCGAGACTGT CCGCTATGCG TCGAGGCTG TTACCGGCA 120  
CGTCTGTG ATGCATTTC AGCGGAGAA GCTGATGCT GTGACTGAT ATATCGCCC 180  
GAAACAGCC ATCAGCCAT CATGCTGCG ATCTCTCTCC AGCGGCGCAC AGGAGGAGAT 240  
AGCGCTATC AGCTTCTCC GCGGGAGAT AGCAGAGTT TTGAGGACA ACCGATGAT 300  
AGCGCTGCG CAGATGTGG CTCTGATGCG AGAGAGCAG CTCTTATTT CGACACGAC 360  
TCCGGAACA CAGATGCT ATGAGTGT TCCCAACA GTCTCTGAA CCGCTCTGTC 420  
GAGATTCOA ATGACAAA TCTGCTGCC CTTTTGTGCG GCGACAGAT GCTGCTGTC 480  
AGTAGAGGC CGAGGTCAA GAGATGTGA CGATCTTAA GGAATGTG CATCTGTGCC 540  
GCTGTAGAT GGTGTGATG ATGAGACAT CCGTAGGAG CAGGGCTTTA TCAACTACT 600  
CAGCTGCCC AGCTGCGCC TGTGTGCGG GAGCTTTGA GAGGGCTCA CTTGCTCTAC 660  
AGCGAGACC CACTCTCTG TCGAGACA GCGGCTGAG CTGACACC TGTGTGACA 720  
GTACATCAGA GAGCAGGCG ABAAGATTT CTGTATGTC GCGGATGCG AGCGGATC 780

CTGAGACTG TCGAGCTG TAGGAGCT GTTTAGGAG CCGTGGGAT AATATCACT 840  
TGGTTATTT GCTGTGCTT CCGATGCG AGATGTGAA GAATGTGGG TCGGGGAGT 900  
TGTGTCTAC TTGTTTCTA CTATATATA TATTTGAG TATATGCGA CTTGAGATG 960  
CAGAGATTC CATTTGAT GTCACTACC GCGTGTGCT TTATTTTTCC CAATTTGGA 1020  
CGTATAGAA GCAAGATTC TCGATTTCC AGGTCCAGG CAGAGTCTT GAAAGATTA 1080  
GCTATTTTC CCGTGGCTC TTGGTCACT GCTGTGCTG CAGTGGCTC TGAAGGACC 1140  
CGCTTGGGC ACAGCTGCG ACTGACAG TAGCTCAAC AGCAGTTCT GCTGAGATG 1200  
GAGCTGTG AGAGCTGCT GTTGGCAGG CTTTGTGCT ATGCTGTTA CATGTATTA 1260  
TTGCTTACT GGTGACACA TTGTACCAT TTCAAGAGA AGGAGCAGG AATTTAAGT 1320  
GTTGTCTAA GGTATGCG TTATATGTC GAGAGCAGG GACTTTCAAC AAGCGCTCT 1380  
CTCTGAGAC CCGCTCTGA ATTCTCTAC TAGGCTTCC TCATCACTT ACCGAGAT 1440  
GCGTCCATC CAGCATGAG GTTGTGCTG ATGTTATTC TCGAGCTG AGGTGAGC 1500  
TGTGAAGAT TTGGAGCAC TCGTTATTA TAAATGAAA TATATCTTA AAAAAAAAA 1560  
AAAAAAAA 1569

## (2) INFORMATION FOR SEQ ID NO: 46:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1924 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 46:

GGGGCCCCC HCGKTTTTT TTTTTTTTT TTTATTTGG ATATGCTTT TATTACGAG 60  
AATGAAAGT TCAATCTCC TTCACTGCT TCTGTTGAT TTCTGGACA CAGCTCACT 120  
GATCTGCTA GAAAGTTGT CAGTCTGCT GTGCTTCCC TCTGTATTT ACTCACCTG 180  
TGTGATGCT TGAAGATAT CTATCACTT CATGTGATG AGCACTCCA TATCAGCCA 240  
CATCATCAT TCTTACCTTA AGATATATA GAAATGTTA ATATTAAGA CAAAGGTATA 300  
AATTAAGST TTGAATGTC TAGTCACTT CAAATTTTA AGATTAANA TCCAGAGTA 360  
AGATTTGGG GTATGTTACA GATTAANA ATAGAGAGA ACTTCATGT GGGGGGGAAA 420  
TCTAATTTA TTCTTACATA AATTAATAG ACACCTGAT TAGAATGAA ACTGTATTT 480  
CTTTAATG TAAAGGCTG ACTCTGAT TCAAGCTCT GAGCAGAT TTGACTGCA 540

CCGAAATAT ACTATCCCTT ATGTGAAGGT ATGTGACAC GTTGAAGCTA CCAATGAGT 600  
TTTAACTCA GCTCTTTTTC CATATGAGC GAGTAAAGCT GCTCCCAATT CAGATATGTC 660  
5 TTTCAATGTC AGGCAAGGCTG ACGACTTTC GAGAGATGTC TCGAGAGTCC CCAACTGCC 720  
CTTCCACAG TACACACGC TGTAGTCTTT CTCTGCAAT CTTTGTATTT TACTCTATTC 780  
TTTCCATCT AATGCTCTAC TGAATTTTAA AGTTAGAGGT GGAAGAGCTA TCCCTTACTG 840  
GAGACGAG GAAACATTT TTTTCTGAGG GAGAGAGAT TCACTTTCAC TATATGCTG 900  
GAGGGGAC AGTGCACAA ACAAGATCA GCTTTCATTC AAGTTCCAGG TTTTCTCTCC 960  
10 TCCCTGAATG ATTAATGCA AGGATATG ATGTAGAGT TCCCTGTGTC AGATGATCA 1020  
TCAATAGG ATATATATC GTTTTGCAAT CTTCGCCCA TTCTCCAGAT TGTCTATCT 1080  
TAACTCAG CCGTTTTCAC TGTCAAAA AAAAAAAA TATTTTTTC AGCACTGGTG 1140  
20 TTAACAGCA AGTTTATAT GTTATAGGT TTACAGCAA CTGTTGAGAT TTCCAGTTGA 1200  
GTCTAAAA TTGCAATCA TTATCTAGA GCATGACAG ATGATTTAGA GCAATCAAT 1260  
CGCTGCAAT CTTCCTCTTA TAGGACGCA TTGTGAGCT GCACTAGCTG ACATCACTAA 1320  
AACATATCA CCAATAGCA AACCAATA AAGGCCAG CCAATATCT GGTCTCTTAA 1380  
AAGCTGCA AATATCTTAA GTGAGGCA GTAAATGAG GACTGTGAGT CATGTACTG 1440  
CAGCTGACA TGAATTACA TAGAGAGAT GCAAGGCCA TTAAAGTTAA AAGTCCAAA 1500  
CTAGTCAC GCACTCTTT ATTTGGGAAA AGTGAGACT TTATGCAATC TTGATAGGTT 1560  
35 TGCAGCTTG CATGAAGGC ACCAATGCA TTTCTTTTAT CTTTGAGAA GCAAGGTAT 1620  
CTGTTCAG GGCCTACAG TAGGAATTA CATCTGGA TCACTCTCT GAACCGAGA 1680  
CTGTCTCAT TAAATATAT TTTGTTTCT AACAAATTA TGAATTACA TGCAGCAC 1740  
TGGATAGC ATATTAATG AAGCACTTA ATTCGAGCT TTGTGAGTT TTTTAAAAA 1800  
CCCACTGAC TAAGATTCAC AATCATTCG TACATACAA TTAAAGCTAG TAGAGACCA 1860  
40 CTAACTGAC AGTTTCTCA TTTTAAATG CAAGAGCTA ATCATCTGA AGTGAACAG 1920  
GTAA 1924

(2) INFORMATION FOR SEQ ID NO: 47:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 47:

TTGTGTGGGG CCGAAGAAC AGGAGACAG TGAAGAACAC CCGAGAGCT TGTATCCGCG 60  
AGGAAGGCA TTGCCATTC TGAAGCTTGG AGGCGAAG AGGAAACAG TGTATCCGGA 120  
5 GGCACATGAG AACTGCTGTC ATGAGAGAG GGCACCTTGG TAGAGAGAT CATTAATACC 180  
AGAACTGTGG TCCCAATGTC CTGTGTGTC TCTTTGAGA AACCAACAG ATACATCTGC 240  
10 TCGAGACCA GTTGGGACA GAGAGAGGTC GAGAGAGAT CTGAGAGAG AAATGAGAA 300  
TAACAGCAC AGTGTATTC ATTCTGTAAA ATTCTATGT AGAGGCTCA GTTGTAGAAA 360  
15 TAAAGTATTT CTACTAGTGG CAGTTAAGT GTTCTGTGTT GTTCTGCTTT CCTGTTAGCA 420  
TAAGTAAACT CCGTTTGGAA CTACAGAGT ATGTCTCTCC TTCAAGATGT GTGAA 475

(2) INFORMATION FOR SEQ ID NO: 48:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 346 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 48:

AGGAGACAGA GACTGTGAT CAGATCTCAT TTACAAATGA AGACCCCAAT CGAGAAATGC 60  
ATGTCTGAAA TTCTGAGCTT ACTGTCTGTC CTGCTGGAC CTGCTCTGGA TAGAGAGAG 120  
35 GAGGAAGAG ACTAATCAGA GGAACCAATG AAGTCACTCC ATGATTTTCC TGAACCTGTC 180  
CGAGTAGAG ATTACGTTT GACHTGAC GTAGAGACT GTGAGATGG CTACTTCTG 240  
GGCAATAGA AGACCAAGC CAGAGCAAG CCGCAAGCTT GCTAAACAG CAGATGCTT 300  
40 GCGAGAGCAA CTCTGTGTA ATCTGCTTC CCGCAGACG TGAACC 346

(2) INFORMATION FOR SEQ ID NO: 49:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1366 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 49:

TAGTGTCTAG CCGCCAGCCC CCCCCCAAT CAGATTTAC TGGCATGAT AGTGCCAGAC 60  
TTCTAACACA GCTGTATTT CAGTCTCTCT GGAAGCTCAC TCAGGAATCA TACCCCTCA 120  
60 GTAGAGCAG CAGGTGATCT TAACTCTTT CAAAGAGCAG CCGTGTCTGG GAAGCATAT 180

5 CTTGACAG CACAGCACC CCTCTGGAA TGGATGACA ACTGACTTCT CAGCGAGCA 240  
GCCAGCTT CTATTTGAC AGTAGGACA GTATAGTGG ATCATACAT CAGCTGGTT 300  
TTTGGTTT TCACTTAG TGTCTGGAC TAAAGTCTT TCAAGTCTTC TTGCGCTTG 360  
AGTGGTGA CCTCCGACC GGAATTCCT CAGTTGCTCT GAGGCTCATG TCTCTCTGG 420  
TGTGCGCCA GCGCCCTGCA TGGGAGGGA CCTCTCTGG GGGGAGGCA GCTGCGGTG 480  
CTCACTTAG GCGAATGAA GTATTTGAG GACTGTTTGT TGAATTTGCT GAGGTTATCC 540  
TTGATGCCA GCGGAGTGC GCGAGTCA CCGAGTGTCT GCGAGGCTTC TCTCTGGAC 600  
TTTCTTCTCT GTCCAGGGG ACCGTGGAGA AGCTGTGAG GCGCGCTGAC TCGAGCAGC 660  
TCTCTGCTG CCTTCCCTGG CAGTGTCTG GGGTGGATTT CCTTACAGCT AGATTTTCA 720  
GCGCTTACT TTCTCTCCAC AAGAGATGCG CAGCCAGCT AGCTCTGACT TCCCACTGCG 780  
ACAAATTTCA GTTAGAGCT CTAGCGAAG ACTGGGCAAT TCGCGAGAG AGAGGAGCT 840  
GTGAGTCTGA CCGVAGSGG GRCGCTTCA CTTTGGCTGG CCTGCTCTG GTCTTAGGT 900  
TTTCTGAGT TGTCTTGT TGTATCTT TTAAGTCTT CAAATTAATA GCGCGTGGG ATGTTGTAG 1020  
CGCGCTTGG AGTGTCTT TTAAGTCTT CAAATTAATA GCGCGTGGG ATGTTGTAG 1080  
AGATTAAGCA GATATGCT GTTTAGAC CAGAGCTTGG GACCAAGGCT CCTTACACTA 1140  
ATTCTCTC CTGATGCT ACAAAGCT TAAATGCT TAACTAGCT TAACTAGCT ACGGCTGC 1200  
GTGAGCAGA GTTCTGAGG CAGCTCTTC AGTGTCTT GTTGAATTT GCTCTGAGT 1260  
TCTATGAA CCGAGAGCC TTAGGAGGC TGGCGAGCA ACGAAGCT CTTAGGCTG 1320  
AGCTGTGAC ACGAGGCT TGTGTCTG TCTGTCTTCT TCTGTCTTCT TCTGTCTTCT 1366  
40 ATAAAGCTA TTCTTATTT ATAAAGCTA ATAAAGCTA ATAAAGCTA ATAAAGCTA

(1) INFORMATION FOR SEQ ID NO: 50:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 50:

55 GCGATATTC CTATTAGCA CTGATGAC CAAATGATG TATTTTTTC CCTGATATC 60  
ATGATTTTAT CTTCTTTTAC AAGGAGATTT TGTCTGATG CTTCTTGGT CCGACTGTC 120  
CAATTATGCG TATATGATTC CATTTAGGG CCGAGGCTTC TCTTGAAGA TGTGTGTC 180

5 TTTAGCCCTT GCGATGATTA ATGAAATTTT GGTGAATGAA AGAAGCCAAA TAGCCAGAT 240  
AGTCCGCCA GCGCTGATA TCCATTAAGG GCTTGGGAAT GCTATATGTA ATTTCTCTTA 300  
GTCTTTTGT TGTTTTAAA AAAAAAACA AGATGGCTC AGATGGATGC CTACGTAAA 360  
ATGTTTCTTA CTTCTTACT CATTAATTTT CTTTGAATTT AGTATGAAA GGAAGAGGA 420  
GGAAGGAAA TTAATATGCT TCTATGATTT CTTGAGCTC AGTCTGACA TATGAGATA 480  
TAACTATAT TGAATGCA AGATTTGAT CTGAAGAG AGAAGCTTT GACATATTA 540  
TCAATGCTC ATATTAATA TTAAGTAAA CCAATTAATA ACAAATGCA ATATGCTATG 600  
CAGAGGCAA AGGAGCTTAA ACCTTTGTCT TCTTTTCTA CATTAGGAAA ATGATTTT 660  
TTTATTTT TTTAGGGGA CTTATTAAT TATGAGGAG TATATCTTT TGTGACTTA 720  
AGCTTATGAA TTGAGTTTAC AATGAGTTG ATCTAGATG GTTACTTACT TGAAGAGAT 780  
GTTGTTGCT TATGTGACAC GAGCGAGGC CTGCTGGGA ATTAAGAGAG CAGTTTTCAT 840  
GCCAGACA ACTGTGACT TTAGTGGCA GATGGGAGT GATTACAGA GTTCCAAAA 900  
TGTGGGCTT TGGGATTTT CCAAGATGC CAGCTGTCT TGTCTATCT TCTCTTTTC 960  
ACACTCTGG ATGATATTT TGAATGCT GAAATATCT TTAATATTT CCAATGAGC 1020  
GTGCGCAC ATCTTATG AGAGTAAAC CAAATGATC CATTATGTTG AATATATGA 1080  
AACCTGTAC CAGCTTGA KTAAGCAGAA TTTTTCCTAT GTTTGTGAA TGTATATGA 1140  
AATACAGG CTAGGATTT ACTGGAGT TTAGGCTCT GGTATATAC TTAGGTTAT 1200  
TTAATATTT GTATTTTAT TTAATGTTT ATGATGTTT GAAAGAGCA AATATATGAG 1260  
GATGGCTCT TGGCATGG TCTTATTTT ACCCTTTT CTATAGGAA AAGAGCAT 1320  
GTCTTATGT ATTTTAAG TTTTGTGAT AGTTTCTAT TCCATTTTA ATAAAGTTT 1380  
TATTAAGAA AAAAAAAA AAAA 1405

(2) INFORMATION FOR SEQ ID NO: 51:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 51:

55 GCGATTTCT AGGAGCCCA AAAAAAAA AGGAGAAA AAGCCGCA ACCGACCA 60  
AAGCCAAA AAAAAAAA TCCGAAAA CAAAAAAT ATAAAAAG AAGATTTAA 120  
AATTTCAG GATTTACT TTAATATG TTAATATG TTAATATG TTAATATG 180



5 TACCTACCA ACATATCTCT GCGCTCTCTC CTGCTCTCTAT ATATGAGACA TACCCGATTC 240  
TCTGCGCGGG CCCCTCTCTG ATGCTCTCTCC GGTCTCTGCTT CCGCGCTGGG TCTCTCTGGGA 300  
CCCTCCAGAG GTGAGCTGGG CTGATGGCC TGGCTCTCTG GTGCTCTGATG GTTTCTCTCC 360  
CCCTACTCTT TTTTCTTGG TTTATCTCA TTGATTTTTT TCTCTGTTT CTGATTAAC 420  
10 CACCTCTGG GAGAGGATA ATAAACATG TATATTTTTT AAGAGGAAA AAAAAAAAA 480  
AAAAAAAAA GCGCGCGGCC CGAA 504

15

20 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 52:

25 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 52:

30 NAAATATCTT GGCATTTT TTAGAGGA GATTAATCA TTCAATGAG ATAGGCAATA 60  
ACATACAGAG AATGAGACTA TCCAGAAAT GCGAGCGGC ATTGAAACA ACATGATAT 120  
CTCAGGACA GATGATGTA TTCTGTTT GTTAGGCTG GAGCAAGG TCAAGATAG 180  
CAAAAATGG ATACAAAG CATATATG TACCAAGCT AATGGAAT GCTGCGCCG 240  
35 TAGGAAAT GCGATTTTG CTATACATA AGTTCTGAA CTCTGATTA GCTTCGAAA 300  
ACATGTTA GGGGAAAAA AGCGAGCTG TTATGAAAG ATATATCTA TTTCATTAA 360  
AATCAATGT CATGCTCTT ATATGATTC TTTATCTTC AGGAGAGAG CATGCGCTA 420  
40 AAGAAACAC AGCTCAGAG CCTCTGATC CAACCTAGAG GTTACCGAA CACAAACTTA 480  
GCTAGAGGT AGATATCTT ATGCTTTTG GTGAGGCAA ACCATTTGG TTGCTACTTC 540  
45 ACAGGAAT CTCTTACCA AGTCTTATC ATATGATAG TCCAGAGT CTCAGTTGT 600  
TTGCAACT GTGCTAGC TGAATAGG CTGAAGGTT AGTTTATG CTGGAACCT 660  
50 ATTACATT GCGAGTAT GTGCTCTG CACTTAAA AAGCTTGG TCTGCGATG 720  
GCGAAGGC CTGCTGCCC CTCTGCGCA TTCTGCTC GAGGCAAT TCTTAT 777

55

60 (2) INFORMATION FOR SEQ ID NO: 53:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 602 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 53:

5 ATGATACAG TGTATACCC TCAATCTTT GAGGTGGC ATGGAACACT GCTTGTATCA 60  
CTCTGTGAC GGTATTAATC CATATATCCA CAAACACA CATCATCCA TCAACATATA 120  
10 CATGTTTGG GATGAGCAG TCAATGTTT TGAAGGGAG TTTTGTCTT TTTTCTTCT 180  
CATATATCT TTAATTTT GTCAATATC AATCAACAA ACAGAAAT TTTTGGAAA 240  
15 AACTTGCAT AGGCTTTTC TATCAATGC TTTAAATAT AGACTAAATA CACATCTCT 300  
GCGATTTTT TCTTACAGT ACATATCTT TACTGCCAT TTATATTAG CCTGTATTT 360  
TTCTGATTA TATTTACTG TACTTGTAT TTTTATTTA AAGAGGAAA AAAAACTCA 420  
20 AAAAAAGAA AATTAAGTGT AGGCTTCTAT TATATATTA TATATATTT ATATTTGCA 480  
CATTTGAAA TACTGTGAA GTTTATCTC TTGCATATC TTTATACGA AGTATTTAACC 540  
25 CTTAAATA CAAATTA TTTTACAGG TTTGATTTT GGTGTGAAA AGATTAATTT 600  
CA 602

30 (2) INFORMATION FOR SEQ ID NO: 54:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1749 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 54:

40 AGTCACTAC TTGAGCGGC TGGGGGAG TCCGCGCAG ACAGGCGGTG GTTGGGATG 60  
CCTCACTCA GTTTGAGAG GTTCCGATC CAAGGGGTT AAAAAAGG AACCCGATC 120  
45 CCGGACACA CTTCGCGCT CCTTAAGG CACACCGGC TACGCTGG CAGCGCGAC 180  
CACTGTCTA AGTCTGTGT GTTGGGGAC GCGCGATGG GAAAGACATC GCTGTGTGAG 240  
50 GATTAATCC AGGACAGCT CAGCAACAC TACAGTCCA CGTGGGAGT GATTTTCT 300  
CTGAGGATC TCCATGTGT TCACTAGAG ATATGCGGC TTCACTGTG GATATTTCA 360  
GGCAGAGC GCTTCACCT TATGACCA TTGTATATC GGAATGCTC TGCCTGTGT 420  
55 ATATATTTG ACCTTACCA TCCACTACC TTCAACACA CCGAGGATG GAAACAGAC 480  
CTAGACCA AGCTCACT ACCCATGCA GAGCGGTGC CTTGCTCT CTTGCGCAC 540  
60 AATGTATC TGTGCTGT GGCATGAG CCGGACGA TTGACCGTT CATTAAGAG 600

5 AACGTTTCA CAGCTTGAC AGAACATCA GTACAGGAGA ACMAAATAT TAATGAGCTT 660  
ATGAGATCC TCAATGAAA GATCATGAGA AATTCAGAG AGAATATCAT GTCTTTGTCC 720  
5 ACCCAAGGG ACTATCATCA TCTACAAACC AATCTCTCCA GCTGGTCTCG CTGCTAGTAG 780  
TGTTTGGCTT ATTTCATC CAGTTCTGCG GAGCTCTTTT AATCTCTTTC CTTTGGTTTG 840  
10 CCACCTGAC CATTTTATTA ATGCAATTTG AATTTCTGCC TGACTACTGT CCAATAGAGA 900  
GGGCCCATG TCACTAGAAA AAGACACTCG GAACCATGT GCATTTCTCC ATCTCTCGGA 960  
TTAGCTTTTC ACATGTGCT GTCATCATTT AATGCCATTT AATGCCCTCG GTGTAGATC 1020  
15 TTCTCATGAG CCTCATTTT GTATCCCGA APTTTGTGAG AAGGATTAGA AATCAGCAC 1080  
TCCCTTTTAG AGATCATAT TCTACCTAC TTCTGAGCTT ATTTTTCCAT TTGATATCA 1140  
TTGATATCAT GACTTCATAT TGAAGAGAAA ATGATATCAA ATGTCAATTC CCAAAATTTCT 1200  
20 TGTAGGCGT TTTTTCAGAT TCTTTCTGTC TTGGAATGTA AACATCTGAT TCTGATATC 1260  
AAGAGAGGG GTCTGGGAT GTGTGATTT TTGGCTACTA GAATGTATCC AGAATGACT 1320  
25 GTATTTTGA AACTCTTAC GTCAATATTA APTTTCTCTT GTCTTGATAT CAGATATGAT 1380  
CAATTTTTTT GGGCGGGAT GTGGCTCAT GCTGTATATC CCAGACTTG GGGAGGCCAA 1440  
30 GCGAGCGGA TCAATGAG CAGAGATTC GAGACAAAC TGGTCAGAT GCGAAAGACC 1500  
CCTCTACT AAAGATGAAA AATTAAGCA GGGGTATGG CAGTGTCTG TAATCCGAGC 1560  
TACTCTGAG ACTGATGCG GAGATCTGCT TGAGCTGCG AGCGAGAGGT TCGATGTGAC 1620  
35 CAGATATGTC CAGCGCAT TCACTCTGCG TGACAGAGAA GGAATCTGTC TCAAAAAGAA 1680  
AAAAAAGAAA AATCTGAG GGGGGGCGG GTACCAATAT CCGCTGATA GTGATCTGAT 1740  
40 AACTGAAA 1749

## (2) INFORMATION FOR SEQ ID NO: 55:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1896 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 55:

55 AAGAGATGG GCTCTTATTT TTCTGAAA AACAAITTTG AGTTACTAT TTATTCATTA 60  
CATTAATTT CTACAGTGA ACTCATATTT GTCCATAGT GCTTCATGAG GACTCATGCC 120  
CCTCTGTCT ACTGGCTCA AATAGACAT GTAGCTTCA CCCCCGTGT TTGTGTCTAT 180  
60 GGTGGGCTTG TGTATATGG AATATAGCA GGTGTGTCAG GTTGGGAGAC ACAGATGTTT 240

5 TTATATAGCT AGAGCTTTA AAAAGCCAG CAGAAATGAA TTCAATATTT GTTATATGCC 300  
TTTTTTTGA CAGATGTTTG AATTTAATG AATTTAAGG GAAACTCAGA GTACTAGAC 360  
GTTATTTAA AGGAAAAAA TGTCTTCAA TGTCTGTAA TCACAGAGG AGAAATGAC 420  
TTGTTCCTT GATCTGTGAG AGGTCAAGT AACCTGGCC GAGCTGTATT TATTTATTTAT 480  
10 AATATAGTAG TAGCAAGTTA ATACTGCTT CTCTGTGTTT CAGGCATAT ATTACAGCTT 540  
CTTTTGACC GTAAATATCA GAATGATCC TCTTCCGAG GAAATGAGCA GAAGCTTAT 600  
GTTTACAGT GTTTGAAATTT GTGATCTGAA ATACACAAA ATTAAGAAACA TGATTTCTCT 660  
15 AATTTTCAA CTAGAGGAG AGAACTTGT GGAAGAGTTC TTTTATTTTC TTTTATTTT 720  
CTTAAGAGAG GCGAGCCAG GTAGTAACTT AAAATATGTC CCGAGGATA TGAGAGTTGT 780  
20 CTTAGAGGT TAAGAGACAC ACTGTTCAC TGTATGCTT TGGGCTGTAG TGGCCAGGGA 840  
GCTCAACTG ACCGTGCCAT GTTGTTTTGA CTATCTAGA CACAGGATC ATTGTATTTCC 900  
TTGACAGGG TCTCAGACC TCGAGATG TTAAATGAGA GAAGAGCTT CTTTCTGGA 960  
25 TATTCATGTA TAAGAGCCA AAGTAATTTT TCTGAACTTC TCGATCTGT AGAATCTGTC 1020  
AAGGATTTA CAGTATTTT AGCTCTGTC AGCATTTTC CATGAGACT TTCTATCAT 1080  
30 TCACTCTTGA GTTCAGAGT TCCATGAT TGTGAGCAG AATATATCT CTTTAGGCGT 1140  
TGGGATCCA GCTGAGAGCA ATCTCTGCA TTTTATACC GGTGTATGTA CAGATATCAT 1200  
35 TCTTTGTGA TCCATGACT TGAAGAGTT TGGAGAGCTC TTATGCAATA TCAGCTAATA 1260  
GGAATGAAA TCACAGGTCA TTAGGTTGT CATTCAGTAA TTCTCTGAAA GCAAGACCCC 1320  
AAGGAAATA TAGTCTTAT CTTTACTATC CACTCTTAAA TTTAATGTGA ATTCTATACA 1380  
40 TGTATATAGT TGTATTTCTT ATATTTTAT AAAATATTT CATCGGAGT TTAACTTCCA 1440  
CTTCATGCT ATCGATGTC TTGGGCTCCA TCGAGAGCT TCGAAGAAA ACAGGAGGA 1500  
45 ATGCAATTC ATATGAGCC AGATCATAT TTCTGTGAG TGAATATAT ATTCTATCAT 1560  
TGTCTTGA AGTCTGAT TCTTTACTT TCTTTGTGTC ATATATATCT AGGTGCCATC 1620  
ACTGATAT GTGAGTGAC TAGAGAGTC ATATATCAT GTTAAGGTACA GTTAAGGCTTA 1680  
50 ACATTTAGA GTTTATAT TTTTAAAAA CTTTCTTGA ACTCTGGGC CAACATGGT 1740  
GAAGACCCCT CTTCTTACTT AAAATATCC AAATATAGC CAGGGGGGTC GATGGGTGCG 1800  
GTGCTGTTA ATCTCAGCT ACTTGGGGA GGGCTTGAG CAGGAGACA ACTGGCTGCG 1860  
55 AACCCCGGG NCGGCGAGTA GGTTTGCGAG TTGAGT 1896

## (2) INFORMATION FOR SEQ ID NO: 56:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1753 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 56:

10 TCTTTTAAAT ATAGACATTT GTGGGCTCA CACATATAT GAATAGTAC CCTCTAATA 60  
AGAGAAAAA AAMATGAGC GGTCAACTT AGAGCAAT TGTCTATTA AAGCAATGT 120  
15 TATTTCACTA GAAAAATTT AATATCAGG ACTATACAT ACTTCATAC TAGCAATTC 180  
TTTTTAATAT GACACTTAA ACAATCACTG AACTTTGAT CCACATACA CCTTTTAT 240  
TTTCTCTAAA CACTTTGAA GCTTAAGTT CTGAGATCA TGTGCACAT GTGATGGCA 300  
GTAAATATAC AGAGAGATG TTATATGCA ATTAAGGCT GTTTCACCT TTATGACCA 360  
CCTGGGCTGT AGTATCTCT GGGGCGAGG TGGCATATG TTTTACAAA ATATATGAT 420  
25 ATGTCAATG TTTGATGTT TGTGTCATG TTGATTTTT GAACAGCAG TTGACCATC 480  
ATAGAAAGTA TTACTTTCTT TCAATGTTT TTTGCTTAC TGGCTTACA GCTTCTCAG 540  
AATATCATG GCGAGCAGG CATACAGTT TCCATCTTA TAGAGATGA ATTAATTTG 600  
30 TATCTACTGA TACAGATTC TGGTCAAT GAAAAAAT CATTTATCC GTCTTTTAA 660  
TATATGTTA AATATATAT TTATGTGCT GATTTTGA GAACAGCTT GAGAGCACA 720  
35 GTTTTCCAT TACTCTTCT GACCAATG CTGCGACCG TTGCTTCTC TTTCGGAGA 780  
GGAAAGGCTG TGTCAATG GCTACATTC TTCAATACC CAATTTGCA TAGCATAT 840  
AAGTATTTA TTTTATGCT CAGTATATA TTATTTAAT TTTTAGTTAA TGGCTATCT 900  
40 TTGCTCATT ACGAAGAA GCAATCATA GAGATTCAG GATAGTTTG TTTAATTTCT 960  
TCCAGATTA ATGTTTTAC AGTGGCTGC TATTCAGAA AGGTATTTCT CTATACACT 1020  
45 TGTTTTACC TTTGAGACA TTGAGGAAA TTATCATG GTTTGTTGAG ATACGACTT 1080  
GATGGCTG TTTAATGAT TTGCTTCAA AGTGGCTAC TCAAGGCC CTAAAGCTG 1140  
50 TAGAATTTA AAGATTTCA AATTTCTT ATCTCTTCT TAACTTACC ACGAATGAG 1200  
GATTTGATA GCAATGATG GTATGATGA GAAGTTTGA CCAATTTCT TTTTGTGTT 1260  
TTGTGTTGT TTTGATTTG AATTTCTT TATTCCTTT AAGATGTTT ATGTATGAT 1320  
55 GTGAAGTCC TAGCAACT ATGCTCAT ATTCATGTA AGTCTGCTT CACTTTTAC 1380  
AAGTTTTCAG ATCGTCACT GATGATGAT ATTTCTTAA TAGATATG TTAAATATAC 1440  
AATGATTTT TAAAGATG ATGCAATCT GTATTTATG TGGTGTGCT GGTCTTATG 1500

5 GGAGCAATT AATCAATTT CATATGATT TTTCAAGTT TGAATCTAC ACATCTACT 1560  
TTGAAATTT CCTTCATCC TGAATACCA ATAGAGAGG CCAATATAT TCCCTCTTA 1620  
TCTTGAAT TTTCACTACT TTATGTTAAA AGTTGTGAT ATTTGTTAAA ATCTGTGAAA 1680  
GATTAAGAG TGAATTTAA TTAAAAAAA AAAAAAATA AAAAAAATA AAAAAAATA 1740  
10 AAAAAAAGG GGG 1753

(2) INFORMATION FOR SEQ ID NO: 57:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1220 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 57:

25 GCGGAGTTA CTGACGCCG GGTGTTGTC TTGCGGAGG GAGAAAGAT TTTTAAACC 60  
CGAGCCAGG TTCTGCTTAC CCGAGGCCG TGTGTGCGG AGACCTCCG GTGAAGCCAC 120  
CGTATCTG TTGACGAGG AGCGAAMACC TTCACTGAG GACTTGGGG ATTAAGAGA 180  
30 AGGTGATAT ATTAAGTCA AATGATTTG ACAGATNAG AGTGAGATC ACTTCATAT 240  
GAATATGCA ACACATCTCA AGAATCTCA AATATCATAC TGTCAAGAC AGGTGTTCC 300  
AATGATTTCA CTGAGTTTC TCTTTGAGG TCAGAGATTT GCTGATATC ATACTCCAA 360  
35 AGACTGCGA ATGAGGAGG AAGATGTAAT TGAATTTTAT CAGGAACAAA CGGGGGTCA 420  
TTCAACAGTT TAGATATCT TTTTATTTT TTTCTTTTC CTCAATCTT TTTTATTTT 480  
40 AAAAAATGT CTTTGTGAT GTGTGTTCA AATCGAATTT GAAATCTGC ACCCATCTC 540  
TTTGAAACAT CTGATATTT GATTTCTAGT GCTCATATTT CATTTATGTT TGTTTTCAT 600  
GTGCTGATTT TTGCTATCA AGCTCATGC CCTTCATAT TACCTCTCC TTTTAAAAA 660  
45 TTACCTGTC ACGAGAGGT CACCTTTTTC AGGACATGTC ATTTCAGGC TTGTGTGAT 720  
AATATATTC GACCATGCA AGTGTTCATA ATGACTTTTC ATTTGGGCTT GATGTTCTAG 780  
50 CATGTGATA CTCACTCCT GCACTGTCAC TTTCAAGTGG AGATGAAAT TTTTCAGACA 840  
ACTGACTGT GAAAAATCA CTTTCTCTTA ACTGAGGCT ACTTTTAAA TTTCAGGGTTC 900  
TGGACAAA GAAGCAAT ATCAAGTTGA AGTCAGATG ACAGATNAG TGAGATATAT 960  
55 GACTTATCC AAGATGCTT TCACTGAGA AAGGCAATTT TAAGATTTT TAAAAATCTT 1020  
GTCAAGAT CCCAAGAG TTCTATTTT CATTAAGCAT TAATTAAGCT ATACATCCAG 1080  
60 AATGATATC ACGAGACAC TGTCTTTT GTTTTATTT GTACTTTTTC GCTTGGATA 1140

TOGGTTTAA ATGACATG TCCTGACAG CTTCTTAAA ATAAACATA TTCTTAAAA 1200  
TCMAAAAA AAAAAAAAAA 1220

5

## (2) INFORMATION FOR SEQ ID NO: 58:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1049 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

TCGCCCTGC AGACACACA TCCTACTGC GTGGGTACC TCCTGTACA TCACTACTG 60  
CAAGCTCCC TCATTTCTG GTGAGCCA TCAGGAGCC ACAGCCCTG GAGAGTGT 120  
GCGATCTTG GCGATGGG AAATGTGCA GAGACAGAC CCGCAATGA GAGCACTAC 180  
CGAGCACCA AGAGTAGCA TTCTGTACA GAGCTTCTC AATAGGGCC ATGCTCTCC 240  
CCAGGGGT CTGGCCCC GCGACACA GCGAGGTGCC AGGCTGGGT CTCTCATG 300  
CCCTTCAT GACCTCAC GCGAGTGT GAGATGGG TTTCGCAAT GGCATCTCG 360  
CAACCTGT GTGAGCGCG TGACTTCAT CTTCTCTTC TTCTGTCTA TGAATCTTG 420  
TTTCTGTGG CTCTCTCG TTGGCTGT CTACTGTG TCCTACTGA GTCCGGGT 480  
ACCTCTGAG GCTGATGG GTGGTTTGT TGAAGGGAC TTGCTGGCC TTGTTGTAG 540  
AGCAGCATA TTTCAGGG ATCTGTGT GCTTCAAGG TATGATAGA GAGGGACCA 600  
CAGGTGTG TTTCCTTT GTGTTAAG GTAGGGAG GAGAGCTTA GTCCACAT 660  
TTCCAAAT GTGGTGGT CCGTGTTC CCGATACT TTTAGTGT ATCGGGCTG 720  
CATTAATGG CAGAAATCA GAGCAGAAA CCGATGGCT TCCCATTT CTCATCTT 780  
TTATCCGAG AGATCTCAG CTGATGCA ACATGTTCC ATGCTGTGG AGACATGC 840  
GAGCTCTCT CTGCTTAGG ACGAGACT GTGGTAGGG CAGGTGAAA AAATTCAGA 900  
CTTTTTAG ACTGTTTT TTATATGT ATATTTTAT TGGCTACTT ATTGTTAG 960  
ACAGTGTA GTGGATCT ATTATTG ACCTTTTCA TAAATAGAT TAAATNAAA 1020  
AAAAAAAA AATCTCAG GGGGGCCC 1049

55

## (2) INFORMATION FOR SEQ ID NO: 59:

## (1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1776 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

AAAGAGAT TGAATAGA GTTCCCAT GCTGTCTGG ATGGAAGA CAGGCTGAG 60  
GAGTCGAT TTAAAGGAC TAAATGTT TCCTGTGAG ATCAGATG GGGACTCT 120  
AGCTTCAT CTTCTTCT TCAAGTCT GACATCTCA GCTCAATG CCACTACT 180  
GTCAATGAT GAGTCCAAA GTTAAGTGA CAGGCGAAA CAGAGGAA TTCTGTCT 240  
TAGCCATG GGCAGAGG GCAAGCTG CCACTCAT CCAATGAT CTAAGGCTC 300  
CTGTGTCTA CTTTCTGA GACTGTGG ACATGCCAA TTTTAGAG CTGCTTGA 360  
GTGACTTTC CTCACTTC CAGCTCTCA CAGTCTTTC TTATGACA TACGCTACT 420  
ACTTAGTCA AGCCCGAT CTTCCTCAC TACAGAGC TCAAGAGAT AAGCTTCA 480  
ACCTCTCAT TGTACCTG GCTCTAAG TAAATGAT CCAATATCA GTTCTGTG 540  
AGGCTCTTC CTGCTGAT GTGGGAGC TGAAGACT TGTATTTAG CTTGTGTG 600  
CTGAGTGT TGTGTGCT CTGAGCAG GCAACAGG GCTCAAGTT GACTACGA 660  
TGGGCGGCA CATCCAGCC CAGGACTCA GTGCTTTC CCGAGCTTC AATGAAAC 720  
ATTAAAGTA GAGGAGAC AGAGACCA GCACTCTC AGAGCTCT GCAAGCTG 780  
ACTGAGTCA GGAACAGC TCTGTGAC TACAGGCC ASCAGCAG AAGCTCTA 840  
AGGCAAGG GCTCCAGG AAGCCAGA TTTGTCTCA GTCAATTA AGTACTTC 900  
GTTTCTCT TGGGATGT GGTTCAGC TGCCTGCT CTTCTTACA GTCTCTAG 960  
AGGCTCTCT GCGCTGCG AGCTGATAT CTAGGTTC TGAAGCTTA CTTCCCTT 1020  
CCCAAAAT AGATCAAC TTCTCTAG AGGATKAAA TTTAGTCT GTTTTCTG 1080  
GTACTTCT TTTTGTGA CTTCTGAT TCAATGCT CCGCTGCA TGTCTCTC 1140  
CTTGTCTCT TAAAGCTCA GATCTGCT CTTTCAAT CATGTGAT AGTAGTGG 1200  
TAGGCTGAT GGGGTGCT CTGTGTGAG CATTAACCA AGGCTTTTT TCTGCAAC 1260  
CATCCCTCA TGCCTGAT CCAATCTTA TACCTTACC CTGACTAT GAGAGCTC 1320  
TGAAGACA TAGGGCCC ACCTTTACT ACAGCTGAG AATCTGGA GCAATCTC 1380  
CATCCAGCA GTCACTGAC ATGTCTAT CCAATCTC TGAATCTG TCAATTTCT 1440  
TTCTCTCT TGGGCTTGG GTCTGGA TGTGCTGT TCAAGCCAG AGCTTAGA 1500  
TGGAGGCT TTCTTACT GTTAGAGT GATCTTTCT TGGGCTTGC CATCTGGA 1560  
AGCTTGTG CAATCTGA AGGTTTAT CTTCTTTCT GAGTTGTT GGAAGGAA 1620

TTGGTTTATA ATGCACATG TCCTGACAG CTCCTATTAA ATAAACATTA TTTCTTAAAA 1200  
TCMAAAAAA AAAAAAAAAA 1220

5

## (2) INFORMATION FOR SEQ ID NO: 58:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1049 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 58:

20 TGGGCGCTGC AGACACAGCA TCTACTGAC GTGGGTACCC TCTGTGAACTA TCACTGACTG 60  
CAGGCTCTCC TCATTTCCTG GTCCAGCCCA TCAGGGACCC ACAGCGCCCTG GAGAGATGGT 120  
GGGATCTTGG GCGAATGGGG AATTCCTGCA GACACGACAC CCCCAGATGA GACACACTAC 180  
CAGCGACCCA AGAGGTAGCA TTCTCTGACA GAGCTTCTTC AATAAGGGGCC ATGCTGCTCC 240  
CCGAGGGGCT CTTGGCGCCC GCGACAGCA GCGAGGTGCG AGGCTGGGTTG TGTCTCATGT 300  
CCCTTCATAT GACTCTACCC GCGAGCTGGT GACATATGGG TTTTCGCGAT GGCATCTCGG 360  
CAACCATGCT GTGGAGCGGG TGAAGCTCAT CCTGCTCTTC TTCTCTGCTCA TGAATGCTTGG 420  
TTTCTGTTGG CTCTCTCTCG TTGGGCTTTGT CTAGCTGGTG TCCCACTCTGA GTCAAGCGGTG 480  
ACCTCTCTAG GCTCATAGGG GTGGGTTTGT TGAAGAGGAC TTCTGTGGCC TTGCTGTGAG 540  
AGCAGGCGATA TTTCGAGGGG ATCTGGTTGGT GCTTTGAGGG TATGATCAGA GAGGGGACCA 600  
CAGGTGTGTG TTTCCTCTTT GTGTTTAGGG TGAAGCAGAG GAGAGACTTA GTCCCAAGAT 660  
TTCCCAAGAT GTGGGTGGGT CGGTGGTTC CCGAGTACTT TTTAGGTGGT ATGGGGGCTG 720  
CATTTAGTGG CACAAATCA GAGCAGAAA CCGATGGCTT TCCCATTTCT CTCATATCTT 780  
TTATGCGGAG AGATCTGAG CTGGATGCCA ACATGTTCCG ATGCTGTGAG AAGACATGCC 840  
GAGCTCTCTT CTGCTTAGGG AGCAGACTT GCGCTTAGGG CAGGTGGAAA AATTTCCAGA 900  
CTTTTTTAC ACTGTTTTTG TTTTATGGT ATATTTTAT TGGCTACTTT ATGTTTAGG 960  
ACAGTGGTA GTGGCATCT ATTTTGTG ACCTTTTCAA TAAATAGATT TAAGTAAAAA 1020  
AAAAAAAAA AAAAAACGAG GGGGGGCC 1049

55

## (2) INFORMATION FOR SEQ ID NO: 59:

- (1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 1776 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 59:

5 AAGAGGATG TGAAGTAAA GGTCCCGCAT GCGTGTGGG ATGGGAAGCA CAAAGCTGAG 60  
10 GGAATCGATT GTAAAGGCAC TAAGTGTTC TGGGTGAGA ATCAGACATG GGGAGCTCT 120  
AGCTTCATAT CTTCTTTCT TGCAGSTCTG GACATCTCTA GCCCAAGTCC CCAACACTCA 180  
GTCCAGTGAT GAGTCCGAA GTGAGGTGA CAGGCGAAA CAGGAGCAA TTTCTGCTCC 240  
15 TAGCCAGATC GGCAGAGGG GAGGCGCTGG CCAACTCAT CATTAGGTTG CTGAGGCCC 300  
CTGCTCTCTA CTTTATTGA GAACTGCTGG ACATGCCAA TTTTAGAGAG CTGCTGAGA 360  
20 GTGACTTTGG CTCTACTTTC GCGTGTCTCA CAGTGTTCG TTATGGGACA TACCTGACT 420  
ACTTAGCTGA AGCCCGGAT CTTCTCTCAC TAAAGAGGC TCAGAGAGT AAGCTTGCAC 480  
ACCTCTCAGT TGTCAACCTG GCTGCTAAG TAAATGTAT CCAATATGA GTTCTGTGG 540  
25 AGGCTCTTGC CTTGCTTAT GTCCGCGAC TCGAAGACT TGTGATTGAG GCTGTGTATG 600  
CTGAGCTGCT TGTGTGCTCC CTGAGCAGG GACACAGGG GCTGAGGTT GACTTACAGA 660  
30 TGGGGGCGGA CATCAGGCG CAGGACTCTA GTGCCATTC GCGAACCTK AAAAAAAC 720  
ATTAAAGTTA CGAGGCGAG AGCAGCCGA GCGACTCTC AGGACTCTGA GCAACACTCG 780  
ACTGAGCTGA GGAACAGGC TCTTGGCAC AACAGCCC ASSCAGCAG AAGCCTCA 840  
35 AGGCGAGGG GCTCCGAGG AKGCCAGGA TTTGTGCAA GTCCAAATGA AAGTACTCTC 900  
GTTTCTCTCC TGGGATGTT GGTTCGCGG TGCCTGCTG CTTCTTAGGA GTCTCCAGAG 960  
40 AGCTCTCTGT GCGCTTGGC AGCTGTTAT CTAGGTTCA TGAACCTTCA CTTCCCTTAA 1020  
CCCAAGACT AATCAKACC TTCTTAGGG AGGAGKCAA TGTAGGTCTAT GTTTTGTG 1080  
45 GTACTTTCTG TTTTATTGTA CTTCATGTT TCCATGCTC CCGCTGCCA TGCCTCTCTCC 1140  
CTTGTTTCTT TAGAGACTCA CACTCTGCC CTTTCAATTA CATCTCATG AGTAGTGGG 1200  
TAGCCCTCAT GGGGTGCTT CTCTCTGGAG CATTAACCA AGGCTTTTT TCTGCCACCC 1260  
50 CATCCCTGCA TGCCTGATCC CAGTTCTTA TACCTTACC CTGACTTAT GAGCAGCTTC 1320  
TCAGAGCCA TAGGGCCCC ACCTTACTC ACACCTGAG AATCTGGGA GCGAGTCTCC 1380  
CATGCCAGGA GTCACTGGAC ATTGTATCC TAGAATCTG TCNCACTACA GTCATTTCTT 1440  
TTCTCTCTTC TGGCGCTTGG GTCTTGGAAA TGTGTGCTCT TCAACTCCAG AGCTTAAAGA 1500  
TGGCAGCGGT TTCTTAACAT GTTAGAGAT CATCTTTCT TGGCGCTTGC CATCTGGGA 1560  
60 AGCTTGATGG CATCTCGA AGGTTTATAT CTCTTTTGT GAGTTTGGTG GCGAGGGAA 1620

GGGTATATAG ATTGTATATTA AAAAAAAAAAG GTATATATATC ATATATATATC 1680  
ACCGAGAAAT AAATCTATCA GAATCTATC TACAAATTA AAAAAAAAAA 1740  
AGGATTCGA TTTCAGCTT ATCGATACG TCACCC 1776

(2) INFORMATION FOR SEQ ID NO: 60:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 443 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 60:

ACAGTAAT AATATATTA TAAATTAAT TAAATATAA ATCTGAGCTA ATCTGATTA 60  
ATTGAGAGAT TTCAATGAA ACCAGGAT TTCTGGCTTC CAGGAGAGT CAGAGAGCT 120  
AGCTAGCAAC ACTGGCTGCG TTGGCTACT TCTTTGGAC AACTGAAAT CTAGCTCCCT 180  
TTTTTTTTT TTTTGGGCC ACTTCATCA TTCACTATAC CTGGCTGGCC TCTGAGGTA 240  
AGTGAGTAT CAACAAAT GTACGACAG TTTTGTCCCT GAACTAGCTG GTTTCAGGTC 300  
CAGCTCTGCC ACTGTCTAGC ATGACCTGGT GCGAAATTC NGCAGGAGT TTTTTTTTT 360  
TTTTTCAGTG CTCAGTCCC CCTATGAG AATCTGGCC CCCCCTGGA CAGATGTTTC 420  
ACCTGGGCC CCGAGTCCC TCA 443

(2) INFORMATION FOR SEQ ID NO: 61:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2888 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 61:

TTATGTGTTT CATTACAC CAGCGCAAC AGAATTTATA TGACCTGGAT GAAGATGATG 60  
ATGGTATAG TTCCGTTCCT ACTAATACA TGAATTTTGC AGCTCTAGCG GNCCTTCTCC 120  
ACCAATGCG TGGGTATAG AGTTTCAGC TTTCATATGC CAGGCCCCC CTCTCTACCA 180  
AATGCTTCA GAATATGCA TACAGAGTC CTGCTCTCCC TTCTCTATT CTGACAAAG 240  
CTTTGACAAA CATGTCGCG ACACTGTGA ACAGGAGAA ACCCCCCAA GAGCTGCCAG 300  
CTCTGAGCC TGTCTCAG CCATGAGAG GCACAGAT GACTGTGAT AATCTGACC 360

CTCAGTGC TCAGAGGAC ATTGTGAGC TTTTCTGTGT GTGTGGGGC CTCAGCCAG 420  
CTCAGTGT CATCTCTGG GTAGGGGAG TGGTGTGTT GAAAAAGAC GATCCCATCA 480  
CGCATATTA GAATATACAC AACGGTGTG TGGAGGGCA CGCATATGAG TCAAGCTTC 540  
ACATGATGG GAATTTATC AACTGAGCC AGCCATCTT GCTGGGGTG AATGAGAGCC 600  
CATCATGAA AAGAGAGCG GAGCTGGCTC GCAGGTGAA CTCTGGCTCC TCTTCACCC 660  
CCCCCTGTA AATGAGCCCT GACACATCC TGAAGGACT CTGAAATGC TGAAGGGGCT 720  
CTATGACCA CGAGCCACA GAATTCAAA TCAAGCTTTG AGCAGGGGAG TGAAGGAGCC 780  
AGAAATGGGG CGAGAGAGG GTGGCTCTGT TTCCCCNAGS CAAAGCTTAT GACCAATGG 840  
CAATGGACT GGAGAGCCCT GATTTGGGA AGGGTTGCC GGAATTAAGA GCTTCTCTAC 900  
TGAATGGAC CGGCTTTCT GTTGTGTGTT CTGGCTGTG CTCTCTCTC TACGTTAAG 960  
TTTTCTGAG TATGTTTCT CATCTCATCG CCAGGTAGS CTGTGTGTTT TCAATGTG 1020  
CTTCCCCAG CTCAGCCCC AAGCTGAAT CTATCTGGA AATGTACAC TGAATTTCT 1080  
GGTGGGCTT CTATGGGCC CATGGATGC AGCCTGGGG CTGTCTGAG GAGCTCTCT 1140  
TTTTCAGGG CCGAGGGCT GCTTTTCTT TGTGTGAT TAACTTTTCA AACATGAG 1200  
GGATAGGGA GCGCTGTGT CTCAGAGGGA GCGAGTGGG CTTGAGAGCT GTGGCTGTC 1260  
TCTGTCTGT CAGTGAAGT GCTTGGGTGG GAGAGAGTC TGAAGGCTCT TATCTCTCC 1320  
CGAGTGGTC CAGGCTCAC TAAATGGAG GCGAGATGA GCTCTGAGC CTGGAGAG 1380  
TCTATCTAG TTCTTGGCT GAGTGGGCT GTCTCTCTC CCGAGAGGA GTTCTCCAG 1440  
GTTCAAGCTT GCTTTTCTT TGAATGTGTC TGAATGCCC ACCCGAGCTC TCTTTCCCT 1500  
CTGGGTGCT TTGCTGGAG GGGGCTGTGT TGTGAGGCT CCGGTTCAC ACCTGGCTG 1560  
GCACTTAAC ACAAGCTGT TTTGTGAGC GCGAGCTCT CTTCAGGTG GCGCTTTGAA 1620  
AGGCTCAGC TCCCATGTG CAGTGTCTTG GTTTGAGCT TATTTGAATG GAGAGGTCA 1680  
GTTTGTGCT GGTCTGCTAT TTCTGGCTC AGTTGTCTAC AGGAGATGAG TCAAGGATGC 1740  
CTGAGGCAAT ATATTCAGCT GCGACAGAG GGCATCTGTT GTTCCACTT ATGTGATGA 1800  
CGCCATCAT CCATGAGCAG AGAATTAAT TCTGTGCTTG CGAGAGAGG AGGATCTAG 1860  
GGAGCAGGC AGCTTACCA GCGAGGTGT TTCCCGACA TAGGGGCGAG CAGTTGGAC 1920  
GAATCTTAG AGCGCAGCA GTTCTGTAT GACGAGCCA GTTGTCTAC TGAATGTCT 1980  
CTTCTGTGTT GCGATGAG AGAATCAGC CTGGCAGAG TGAAGGCAAT TGGGGAGAC 2040  
GGTTCTGGA GCTCTGAAA ATCTATGCA AGTCTGGA AGGCGCATG CCGAGATAC 2100  
TGAAGGCTC CCAATTTG CTTCAGGCTA GCGCATGAA AAAAACCAG GACTCATGTT 2160

2220 TCACGGGTC AGTCTGTCTAG CAGAGAGCA CCGAGATTTT GAGCCGAGCT TCACTGTCTCA  
5 GGCTCTGAGG CTGCGCAGCA CGGGAAGTC CAGAGAGGG GCTCTGTCTT GCTCCACTTG  
2280 CAGTCTCTTA AGAGATGCTG CTTTATATTC TCTTACCTCT TTCACTGGG TCGAGACTTC  
2340 TCGTTAGGAG CTGAGAGCA TTCTCTCCAC ACTTTTCTCT TCTCTGCCCA AGAGAGCATC  
2400 CAGAGGCGAG TAGGAGCTGG TTTTTCAGCT ACTCTGAGCC GGGGGCTGAC TCGTTTCACT  
10 GTGCTTAGGG TAGGAGCTGT AAGATCTCTC CCTGCATGGC TTATATCTCTC GTCTCATGCC  
2460 AAGACAGGTA TCTTCTGTCT GTACAGAGT TTCAATGAGT CCAAGCTGAG CCACTGTGGC  
2520 ATCTGAGGCT GATCTGATAG GTACCATCTT GATACATTTA GGGGAGCTGT TTGCTCTCTC  
2580 CACTCTATTA CCACTCATCT TCGAGAGCG GAGAGAGAG GTGTTGGGCT AGTCTCTGTCT  
2640 CTTCTCTGAG TTCCCATGCC TCTATGTTAC CCACTCTGTCT CTCTCTGTCA GAGAGAGAG  
2700 AAGGGGCAAT AAGAGATGAA GGGTGAAT GTATTACTTA TCCATTTCTG ATTAAGACTT  
2760 TGTATATCTT AAAAAAAAAA AAAAAAAAAA CCAAGGGGGG CCGGAGACC AATATGCCSK  
2820 AAGATGAG  
2880

## (2) INFORMATION FOR SEQ ID NO: 62:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1851 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 62:

40 CACTAGATTA ATTATATAT ATTAACTAT CTGAATTTCT TTCAATATAT AGGTGTCTTA 60  
GTTCCTATG AGGTTTGGC ACTTCACTT GCACTGTCTT CCAACTTTTG GACTGAGCTA 120  
45 TCGCAGACT AGTCTGCTAT GTCAAAAAC TGCATCAGC TTTTGTGTGA AGATCTGTCT 180  
TTCCAGAT ATATATATG TATCTATG GATCAAGAA CTTTATTTAA CAGACATAT 240  
50 GTTACACTT TCAATGACA TTCTTTCTTA AAGGTTTCAA GTCAATGTT TTCTGAGCA 300  
AACTGTCCA ATTGATGAG CACTCTTAT ACAAATTTGA TAAGCAGTA TCGAGACTTA 360  
CACTCTGATT TCTCCAGCG AGTTGAATTT TCCAAAGCA GTCTTCTTT AATGTGGAC 420  
55 CTGAGGGCAC TCGTTTCTCT CTTCTGATG CAGACTTCCA AAGATTTAA CCGAGCTTA 480  
ATTCCAGCT TCGAGAGCT TTAAAGCAA TCGAGACTT GTCTGCAGCA GAGAACTCA 540  
60 CTCAGAGCG AAGAGGCAA AAGAGGAAA ACTAAGATG ATGAGAGAG AACTTCCATT 600

AAAGGGGCG GTTTTAGGAG TCAATGAGAG CACACTGTAG ACAGCTGCAT CAGTGCAGTG 660  
AAGAGGAAA CCAAGGAGGT CCTGACCCA AGAGACTT CTGACATGTA GACGAGAGAC 720  
5 TCTCTATTTA TTGATTCAGG AACTGAGCAA GATCTTCTTT CCGCTGAAA TAGTTCTGTT 780  
AAGATPACC GAATGCAAGT TCAATCTTGG TTTTTCAGAG ACATGTCAA TATCAGGTCA 840  
CAGCTTCAG AAGACAGCTC CAGCATGCTT AGATATGAG ATTGTAAAG AATTAAAGAC 900  
CTCCACTGTT CCAAGGATTC TACCTTAGCC GAGGAGAT CTGAGTTGCC TTTTACTTCT 960  
ATCTCTCAG TTCTGTCTGA CTTAGCTGAC TTGAGAGCT GTCATGCCA AGCTTTGCC 1020  
15 TCCAGAGCC CTGAGTTGCG TTATCTCTC AGTTGTGCC ATTTCAGAG AGCTTTTAGT 1080  
CATATGAGC AAGATGAT TTAGATPACC CTGTGTAGA CCAATGATC TACATTCAT 1140  
GTCTGACAA GCAATATCTG CAAAGGAAAC CAGCTCTCTT CTTCACATTA GGTGTAGCAT 1200  
GTCTACTTTT AGTTCCTCTA CCCCAGCC CCACTCTGTT TGTATAGTT TTGCTTATTT 1260  
GTTTTCTGCT TTCAATTTGT CAGTGTCTCT CTGCTGAT GTGAGATAG AATTATAGCC 1320  
25 TTATATCTTG GTTCAGGAGA ACTTCAGATG AAAAAACTT GCACTTTCAG TATACTTCTT 1380  
AAGGGAAT CAGATATGAG ATATGTTTTA TGTATATAG AGTTCACTTT AGTGGCTTTC 1440  
ATTATATG GTCTGTCTGG AAGAGAGGG TTGCTTAGCC CTGTACATG TATTTAAAC 1500  
TTACAGCAT TTACTGTGT ATGATATGCT GTCTCTGTG CCAATTTTGT ACTATPAGA 1560  
GCGAGATTC CTGAGTCC TTGCTTCTTT ATATCAAAA TTAGTTTAC TTGTATAGGG 1620  
35 AACACACA AAGATTTGA TTCTGTAAAG AATCTCTTT AGCTGTGCC TGGCATATA 1680  
TAATGCTGC TTTATTTTAC AGATPACTG TCGAGGAAT AAGGCACT TGTATTAATA 1740  
ATATTTGTT TATTTTAT GACATGACT ATGATGCT ATTCTGTCA CTTAATTA 1800  
CTGATGTGA TCACTTAAA AAAAAAAAAA AAAAAAAAA A 1851

## (2) INFORMATION FOR SEQ ID NO: 63:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 63:

55 TCAATGCTG ATGAGCTCTT TCGTGGCAG GCGACTCTCT TCGTCTGTA TGACTTTTC 60  
AAGAAATTC CCACTATCTT GAAAGATTC AATAGTATC CCGTATCAT TGACCTCTCT 120  
60 GCGAGGCA CAGATTCAT TATGATGAA TATAGGATC GTAGATCAC AGGAGCAGC 180

5 TTCTGTGATG AGCGCTTCAG AAGAGACTTA GAGATGAC TGAGATTGCG TAAACCCCTT 240  
CTGTGTCAGG ATGTGGAAAG CTAGATCCA GTTTTGACC CGGTGTGAG CTGTGAGATG 300  
CGCGAICAG GCGGAGAGGT CTTGATPACT CTGCGGAGCC AGGACATAGA CTTGTGCGCA 360  
TGTATTGTCA TCTTCTGTCT CACCGGAGT CCAACTGTG AGTTCCGACC AGATCTCTGT 420  
TCCCGGGTTA CTTTGTGAAA CTTGACAGT ACCGTGAGCA GTTTTACAG CAGATGTCTA 480  
AATGAAGTAC TTAAAGCAGA AAGACCTGAT GTGAGCAGCA AAGCATCTGA TCTTCTTAAA 540  
CTTCAGGGGG AATTTCAGCT CGTTTGTGCT CAGCTGAAA AATCTCTACT ACAAGCTCTG 600  
AAGCAGAGTGA AAGGCGCAT TTTGTGATGAC GACACAGTCA TAACTACTCT GAGAACTCTG 660  
AAGAGAGAGG CTTGAGAGGT CACAGGAAA GTTTAGGAGA CGGACATTTT CATGCAAGAG 720  
GTGAGAGAGG TGTGCGAGCA GTACTCTCCG CTCTCCAGCG CTTGCAAGAG CATCTACTTC 780  
ACCATGAGGT CCGTCAGCA GATACATCTT TTGTACAGT ACTGCGCTCA GTTTTCTCTG 840  
GACATTTATC ACAGCTGCT ATAGCGAGAC CGGAACCTGA AGGTGTGAC CACACAGCA 900  
CAGCGCTCTT CATTATTAAC AAGGAGCTTC TTGAGGTG GTTTTAAAGG AATTGCTGCA 960  
GGCATGCTGC ATGAGAGCA CATTACTCTT GCATCTGTC TGGAGAGAT CAAACTGAG 1020  
GGCAGCGTGG GGGAGCCGAC CTAGATGCA GATTCGAGC ACTTCTGTGAG AGGAATGAG 1080  
ATTGTCTCTA GTGTGCTGCT CACCGCCAGG ATCCAGGCGC TGACTGTGGA CGAGCGCGAG 1140  
CGGTTGTGTA GCGTGAAGT CTTTCCCGG GTTAAAGACT TGAATGCGAA GGTTCAGGCA 1200  
GACGAGCAAT TTGCAATCTG GTTGAAGC AGCTCCCGG AGCAGACTGT GCGCTACTTC 1260  
TGGATGTGAG AAGACCTGC AAGACCAAT GCGAGGCGCA TGCACCGCTT GCTGCTGATC 1320  
CAGGCTTTCC GCGCGATGCG CTTGTTGCGC ATGCGCCACA TGTTTGTTC AAGAAACCTT 1380  
GGGAGTCTT TCAATGCTAT CATGAGAGG CGGCTGACC TGAACCAAT TGTGAGCA 1440  
GAGGTGAAGC CAAAGCTCC TGTCTTAATG TGTCTGTGTC CTTGTTATGA TGGCAGTGA 1500  
CATGTGAGG ACTTGTGAG CGAGCAGAC AGCGAGTCA CTTGATTTG AATCGGCTCT 1560  
GCGAAGGCT TTAAAGAG AGTAAGGCA ATAAAGCG CTTGTAAGTC GCGCAGGTG 1620  
GTGATGCTGA AGATGTGCA TCTGCGCCA GGTGTGCTGA TGCAGTGA GAGAGATTTG 1680  
CATTCCTGCG AGCGGATGCT CTGCTTCCGA CTTTCTCTCA CCAATGAGAT CAACCCGAG 1740  
GTGCTGTGA ATGTGCTGCG TCGCGCGCG ATCTTTGTGT TGAAGCCACC GCGAGGATG 1800  
AAGGCCAACA TGTGAGGAC GTTCAGGAG ATTCGCGTCT CAGCGATATG CAGTCTCC 1860  
AAGAGGCTG CCGCTTGTTA CTTCTGCTG GCGTGTGTC ATGCAATCAT CAGAAAGCG 1920  
TTAGCATAG CACACTGG GTGTCAAG AAGTATGAT TTGAGAGTC TGACCTCGG 1980

20 TCAATTTGCG ATAGGTGGA CAGTGTGCTG GATGACAGCG CGAAGGCGAG CGAGACATC 2040  
TACCGGATA AGATCCGCTG GTCTGCTCTA AAGACTTAA TGGCCCATTC CATTTATGCG 2100  
GCGCGGCTGG AAGACAGTT TGAAGAGCT CTGCTCAACA CTTTCTGGA CGCGCTGTC 2160  
AAGACAGGA GTTTGAGAG TGAGTTTAAAG CTGCGATGCA AGGTGAGCG AGATTAAGAC 2220  
ATTCAAAATG CAGATGCGAT GCAAGCGAGA CGAGTTTGTG CAGTGTGCTG AGTTGTGCG 2280  
CGACAGCGAG AGCGCTCTCT GCGTGGGCTT GCGCAAGAC GCGGAGAG TGCTGCTTAC 2340  
CAGACAGGT GTGACATGA TCAATAAAT GCTCAAGATG CAGATTTTGG AGATGAGA 2400  
CGACCTGCGC TACCGAGGA CTGAGAGAA GACAGAGACA GATTCAGCT CCGAGCGGCG 2460  
CGCTGCTGAG ATGCGAGAC TGACAGCAC CGCTGTCAAC TGGCTGAGC TCAATCCGCA 2520  
GAGCTGAGC CAGCTCAAG CGACCTGTGA GATATTCAG GATCTTTTGT TCAAGTCTTT 2580  
TGAAGAGAA GTGAAGATG GCGCAAGCT GCTTCAAGAC GTTGGCGAG ACCTTGCAGA 2640  
TGTCTCTCAG GTGTGCGAG GAAAGAGAA CGAGAGCAC TACTTGGGCA CGCTGATCAA 2700  
CGAGTATG AAGGAGTCT TGGTGTGAG CTGTCTCCAC TACAGGTGCT CTGCGCGAT 2760  
GAGCTCTC CAGTGTGCT CCGACTTGA CGAGAGATC AAGAGCTG AGACATCTC 2820  
ACTGCGAGT GATCTGTG GCGCAGGA GCTAAGAGAC ATCCAGTGT GCGTGTGCTG 2880  
CGTGTGCTG CCGTGAAGCT AGATCTGCT CACAGCGAG TATGTGCGC AGGCGAAGAG 2940  
CTGTGCTG GAGGAGCTT GCGTGAAGT CAGGTCAAC ACCTCAAGG CGGCGACCT 3000  
TGAAGCTG AGCTGTGAG TCAAGGTTT GAAGCTTGA GCGGCGAGT GCMAGACA 3060  
CAGGTGCTA CTGTGATG CCAATCAAC GCGCTTCCG CTGAGCGAG TGGCTGTGT 3120  
CAGCGACA AAGACCGAG AGAGCGAG TGTGTGTAAC TTACTGTCTT ACTGTGAAT 3180  
CAGCGTGA GACTCATCT TCAAGTGA CTTGAAATTT GCTAAGAG AGGATCTGCG 3240  
CAGCTTAC GAGCGGCTG TGGAGTCTT GTGACAGAG TAACTTTTC TACTGCGCG 3300  
TTTCTGTAA AGTGAAGTT GATTTTAC ATTTATCTT TTTTAAATA TTTGGAAGT 3360  
CTGAGTCT GAAGAGAG TGTGTGTCT GAGTGTGAG GAGCTGAT GATCTGAC 3420  
GTTGTGAGT GGTGAATTT GGAAGATAC CAGAGATAT TTGCGAGCG CAATGCGTG 3480  
GCTCTTGA GGAATAAA CACTAGCAT GAAGAAAAA AAAAACTTA CAAGCGCAA 3540  
GG 3542



(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 883 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 64:

10 AGGTGATTTT ATGATAGTGT GTGATGATA GACGGATTA TCTGTTTACA TTCTGTTCTT 60  
CTGATATCAC TCACACAGCG GTAACTAGGT GACAGATCTT ATTCAGAGCA 120  
GGTCTTACAG CAAACCAAGG TCTCATCTTC CCAATGATA GATGACGGCG CCTTGAGATA 180  
15 AGCTACAGGC AACACACTTT CCGGTTTCTT CTTGGCCCTT GATTCCAGAT GCGCGATGAA 240  
GCCACCGAC GTTATGATGTC TCAGATCCCG GTGCTGAGCA CTAAACGCCCG ACCCGAGAT 300  
CTGAGTTCTT GGTGTCAGCG ACTGAGGAG GATATACAT CTTTATTCGG GTATGTGAG 360  
AACAACAGCA ATGCTGACAA CAAATGCTTC GAACTGAGT GCACACAGCA AGGAACTCGG 420  
TGTGTTGGAA AATCTGATA TATCATGAC CTCCTGAAAT ATGAGTTTGA CATCGAGTTT 480  
20 GACATCTTCA TCACATATCC TACTACTGCC CAGAAATGTC CAGTTCTTGA CTGAGATGGA 540  
AAGACAGCA AGATGATCAG GGTGGGAAA NTAATCTTGA CCGATCATTT CAAACTTTTG 600  
TGGGCGCAGG AATGTGCCA AATTGGAAT AGCTCATCTC ATGCTCTTGG GCTTGCGTCC 660  
30 ATGCTGGCA GTGGAATCC CTGATCTGAT TCAGAGAGCG GTCAATCCAC ACAAGAGAAA 720  
ATGCACAAA TCAGAGATCA AGCCACTGAG CGAGCGCAGA GCGACTTTTG ATAGGCTAGG 780  
35 ATACTATTTT CTTGTGCATC AACTTAACT CATCTAACTG TTCCCGCGAC AACCTCCACT 840  
CTACTTTTAA CTAAATATG CAATGACAT NTGCGAAGA ACA 883

(2) INFORMATION FOR SEQ ID NO: 65:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1541 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 65:

50 GCGACAGAGT GGCCTTACC CTGGCTCAT CTGGCTACAC AGGGAATCTA AACGCTTCCA 60  
GATTCCTCGG AACATGCCA CCGCGATAG CCTTCACAA GAGAGGAAA ATACATTTT 120  
55 TAAAGCCTGG GCTGTAGAGA CAGGAGATA CAGGAGAGG GTGGATGACC CTGACCCGAG 180  
TAATATGAG GCCGAGCTGC GCTGTGCTCT CATATGAGC AGCAATTTCA ACCTGATATA 240  
60 TCATGGGACC AAGAGATGTC CAAATACCC AATGAGATA TATCAAGTGT GTGACATCCC 300

5 TCAGCCCGAG GGTCTGATCA TTAAACCGAG ATCCACAGGG TTCTGCTCCCT GCGATGAGAA 360  
GGATATGAT GTGATGAGAG AGATGAGAGT GATCATGTGC AGCAACATGT 420  
TCGCATTCAG GACAGCTTCC CTTTCTTGAA CATCAATGAT TTCTCCATGG CCGCAGCCAG 480  
TGTGGCCAT TCAGATGTGG GAACTGCGAG CCGGAGGCA GTGTGGGCA AACATGACC 540  
10 CCTGAGATG GAAGTAAACC AGGCACTTAT ACAGCCCTTC TATAGCTCTC CAGAACTGTG 600  
GATCAGCTCT CTCCCAATGA CTGACCTGGA CATCAAGTTT CATTAACCTG GGAAGAGATA 660  
CGGCGAGACC ATGACCTTGA GAAACCTTCA GGGTGGGCA CTCCTCTATG GCGACCTGGG 720  
TCCATGCTT GACAGAGAG AGCTCTTTGG TCCGTGAGN CTGAGGAGG TCAGATTTCC 780  
AGTTCCTGAG CATATTAACA ATGAGAGCA GAAGCTGTTC ACTAGCAAGC TCGTGGACGT 840  
20 CATGAGAGCA GAACTATCCC TGGAGGTGAG CCGTCAATCC ATTTATGCCA TCAGGCTGTG 900  
CGATGGAG GTTACTGCTT CTGGGCATG TCCGCCATCA CTGTTTCTTC CCAACCTCAT 960  
TGAGAGACA AAGAAATGTA AGCTATTTTG TCTGAAACA TTGCTTAAGG ATCTCATTTCC 1020  
CGACAGAAA GACAGATGAG AAGAGCAGCC ACCGTTTGAAT ATCTACTTAT GCTTTTGCGA 1080  
AGATGGCCA GATGGGAAC CATTTGAAAG GAATCTATC TTGGTTGAGG TCATTCGAGT 1140  
30 AGTGGCTGG ATGATCTAGG AGATGTTTTC TGGTGAATTC ACAGATCTCT TTGATATGAG 1200  
CAATGTCGCG CTCAGATCT CAACCCAGA CATCAAGAT AACATGTTG CTCACGTGAA 1260  
CGAGCTGAC GGAATCTTTC AAACCCAGAA GAUCTGGCAG CCAATGAGC CCAACCCGAG 1320  
35 CATGCACTG CCGCTGCCC TCGTCCGCA GTAAATGTGA ATGCAATCTT CTTCCTTCTC 1380  
TTTTTTTAAA TATGTTAGAT ATGCAATTTT TTAATGTTTA GATTTAACA GCTTTTAAAT 1440  
40 CTCTGTTTTC TGTGACATG TTAGAGATTT GTGATCTCC AATATGCTT AGATTTAAG 1500  
CTGATTTAAT TTATGAGAAA AAAAAAAAA AAAAAAAAAA A 1541

(2) INFORMATION FOR SEQ ID NO: 66:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 732 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 66:

AGAAATGAAA TATTAGAGGG TCGTCTGCCA GCGGAGAGAG AGTGTTTTCT CCGCTGGAG 60  
AAGGCTCTGC TCAGCCCTCA GAGTCCCTTC CTGCCCCACC GATTAATGCCA CTTTAAAGAG 120

5 GAGCTGACC GCGAGCTGTC CAGACGATT GCGCCCGAG AGTGGGGAG TTCTGTCTCG 180  
CCCTCTCTG TCTGCTGAC CTCACCCAG CTAGAGGGA GGTGCAATCA GGTAGATTT 240  
GCTCTCTATT CAAATTTCTG GGGCTTTGGG CGGAAMCAG CAGCTTTGG GCTGTGTGG 300  
GAGACTCTTC CAGACAGGA ACCCGAAG GAGACAGAG CTCGCAATC CTCGACGGC 360  
AGGCTCTGG CAGAGGTAT TGGACTGAG ATTGGGCA ACDCAAATTC ATCTGGCTG 420  
GAGCAGAGG CCGAAGAGC TGGCTTTGTC CCAATGTGG AGCTCTGTC TCAGCCCTCT 480  
TGTCCCTTG AGCTCACTA ATTCCGACA GGTCCGACA GCTCTGAGC TTCAAAATCT 540  
ATATATTGAG AGAGTTGAG AGTATATCAG AGATATTTT GGAAGGAGT TGGCTATTC 600  
ATGTGATT TGGATCTTC TTGAAGTTT AATGTTTTT TTAGAGATT TAAAGAAAT 660  
AAGGCTAC AATATCAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 720  
AAAAAAAA AA 720

25 (2) INFORMATION FOR SEQ ID NO: 67:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 629 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

35 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

TTAAGGAATT CCGCCGATC CCGCAAGTA ACATGACTTA AAGAGAGCG GAGATCTCG 60  
GGTCTCTCT AGAGTCGAT GGGCTAGAG AGAGCTGTC CCAATGTGG AGAGACTCG 120  
AGGCCGTGAA CTCGAGACT CACAGCCGG AGCTGAGCCC AGAGCCGAG AGGTCCCTCG 180  
AGAGGAGAA AAGAGGCTA ATGAAACAG CTCGCACTA CGAGAGGAA CTGAAGTTTC 240  
TTCCGCAAG GAAACCGAG AACATCTGC TCTGTGTGG CATTCTTAT CTCCTGAGC 300  
TGTCTATGC CTACTGAGC ATGTGAGCT GGCATTTCC CACAGCCAG ACAGGCTTCC 360  
ACTTGGCCCC TTGCTAGGA TCAGCAGGC ACTTCAGCC TCATAGGAC CAGGTGCTG 420  
GGGTGTTTCC CTCGAACTT AGTGTGAG CATTGTTTC TGGCGGCCCA GGGCTTGGCT 480  
CCCTGGCTG CTGGGGGTT CCGGCTTCC AGAGGAGT GGTGCTGTC CTCGCTTAG 540  
CCCAAGGGAG AGGCATPAA GAACACAG CTGAAMAAA AAAAAAAAAA AACTGTAGG 600  
GGGGGGGGT ACCCATGCG CATTCTG 629

60

(2) INFORMATION FOR SEQ ID NO: 68:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1751 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 68:

5 CTCTTAGGCG GCGGCGGAG GCTCCGAGC GGTGAGGCG GCAAGCCAG CCAAGGCGT 60  
GGTACCCGCG GGTGCGGCG CTCAGTCTGC GGCATGCGG GGTTCGCGC GGTCTGTGG 120  
AGGGGTGATC ATGGGGGCG CCGGCTCGGG CAGGGGAGC GTGTGTGCG GATCACTAC 180  
ACACTTCGAG CTGAGCGAC TCTCCAGCG GAGCTGCTC GCGGACACA TGTCTCGGG 240  
CAGAGAAAT GGTGTGAG CAAAGGCTTT CATGACCA GGAAGACTCA TCCAGATCA 300  
TGTCACTCT CCGGTGGCG TTCAATGCT GAAATCTC ACCGATTA CTGTCTGTT 360  
GATGTGTTT CAGAGAC TCCAGAGC AGAGGCTTA GATGAGCTT ATCAGATCA 420  
CAGATGATT AACCTGATG TGGCTTTGA GTCATTAAA CAGGCTTTA CTGTCTGCT 480  
GATCTGCC GCGGTGGCG GAGTCTATTA CATGATTC AACCTGCCA AAATGTGGG 540  
CATGATGAG CTGACTGGG AGCTCTCAT TCAGGCTGAG GATGATTAAC CAGAGAGCT 600  
TATCAGAGA CTAAAGCTT ATGAGACA AACHAGCA GTCTGTGAAT ATTACAGAA 660  
AAGGGGTTG CTGGAACAT TCTCGGAGC AAGAACAC AGATTTGCG CTATGTATA 720  
TGTCTCTTA CAATCTNAG TTCCAGAG AACCCAGAA GCTTCAGTTA CTCCATGAG 780  
AGAAATGTT GTAACTATTA ATAGTAGAT GGCAGAGCT CTATGCTTT GATTTAGAA 840  
GCTGCTTTTC CTAGAGCTTC TATATGTAT GATTCCTTG AAATATATAT TACTTTTAT 900  
TCTACTGAT TTATTTTGA TACTAGGAT GTCCAAATG ATTGGATAC TAAGATCAT 960  
GTTTGAAAT CATCTAGTGT GTTGTATCA GTTATCTCA AAAGATCAG CAAATCTCA 1020  
ACCTTTAAA CATCTTTAG ACNAAATTA AAGAGCATT TGTATATAT CTACTTTT 1080  
GTTCTGTTAA TATGTGTTG ATAAAGTTTC CATATTTTC TGAAGAGTT AAAAAAGTT 1140  
ACATGTGAT TGGAGAAAT ACATATCAG AAATTTGCG ATAGATGAT GCCAAAAGG 1200  
ACATTTCCAG CATGTGGA CAGTGTGAG CACTATTTAA AATTCAGAA AGAAGGAGC 1260  
TGGATTTACA GATTTATGT GAGACAAA TTCACTGCTG CTTTACTCT AGGAATGTA 1320  
TATGTAGCC ATATATCTG TATTTATTT GTCTTTAGC ATACTTTGAG TTACTGAGA 1380  
ATTTCATTT TCTATNAG ATGTATCAT TACATATAG AAAATATTA CTTTATATG 1440  
ACTGTTTCC TTGAAATA CTGTGACT GAGGTTATG ATTGTGTCA AAAATTTGCA 1500

TAACTGCTTT TAGAGCACC AAGTTTGAT GAATTTTCA CAAATGTAA TTAAATCTA 1560

TTTTTCACT TAGACTCAG GTTAAGAAAT GTTTTTAGG ATCTACTGC TGGTTTTCT 1620

TTTTGATCA ATGTGTGAT CTGGCTTAT AATATACAG TTATGTATC ATCTCCCCG 1680

CGATTAATA AAAAAAAAAA TCGAGGGGGG GCGCGGTACC CAATTTCTCG 1740

NAATGGGAG T 1751

## (2) INFORMATION FOR SEQ ID NO: 69:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 69:

GGCAGGAT TATGTATTA ATGTTTTG AATGTGAA TATTAGATA TTGTACTAT 60

TTTACCAAC TCAAAATTC CATGGAAA TACTGTGCA TACCAACT ATGTGTGAA 120

AATATCAT GCATATAC ACTGTATCA GACTCTAGTA CAGTTGGCC AATCAGCA 180

CAGTAAATA TTGAACAA AATCTGAC ACAAACAG CAAAGTGG GGTTCAGAA 240

GCTTGACG GTACTAGT GTAGATCT ATGCACAGT GCGAGTGT ATGTGCAAT 300

CCAGAAAA CTCGAGAG CCCAGCTT CACTCTGT TACCAATG CTCGTGTAA 360

GCGAAGTG AAGCTAAG CAGATTAG CTCGTAGGC ATTCCACAC ATACACAAA 420

ATCTGCAA GCATTAGA AATCTGTA CTCCTAGTG TTCTGACCC AGGACACT 480

CCTACTAG CTCATTAA AATAAAA 508

## (2) INFORMATION FOR SEQ ID NO: 70:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 245 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 70:

TACATAGC AAGAGAA TTCCAAAT TCTAAATC TCGAAGAA ATTTTCTCA 60

GATTCAGT TTCTGTGT CTCAGGTG TGTAGGGC TTTTTCCT TTTTGTCTT 120

TCTCTACT CAGCTTCT CTCCTAGT GTCTGTGG CATGATCG TCTCTACTG 180

CGATTTCTT CACTATACA AGAAACTG CAGGTTAAA AAAAAAAAAA 240

NENCG 245

5

## (2) INFORMATION FOR SEQ ID NO: 71:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 361 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 71:

ATGTTCTCA TAGAGTACA CTGTGCTTC TCCAGTATT GCTGAGCTT CATAGTACT 60

CCGACAGCA CCGACATAC ACTACTTAC CTGTGCTT GATCTGAC CAGCATGCT 120

GGAGAGGA CGACTGGG ATGTACCTA AATCTGTTA CAGGAGGG ACTCCAGAG 180

TCAGACAG TAGGACTTC CTCAGAGT GATACATGT CTCCTGTAT CCACTATTT 240

TTTTTTTT TTTTGATA GATTTTACC CTGTTGCC TGGTGGAT GCATGCTGC 300

GATCTGCT CACTGACCC TCTGCTCT CCGGTTTCAA GTGATCTCC TGGTCAAGC 360

T 361

## (2) INFORMATION FOR SEQ ID NO: 72:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 713 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 72:

AGATCTAC AATAGAGAC ACTGTAGTA CATTTGCTC TGTCTAAG ACCGACACA 60

TTGATCAAT TTTGTTGTT GTTATATAT TTTCTGTAA AATTTGTGA AAGTTCTTT 120

TTACTAGAT GATTTTAA TAGCTGGAG TCTTTGAAA CTATTAAGT GTCACTACTT 180

AACAACATA CCTATGTT TTTTGTGTT TTTTATCAC TCAATATAA TCAGGAGAG 240

TTACCCACC ATCTAGCAT TAGATCTTC TTTTATGTT TCTTTAAG ATATGATGT 300

TCCCATACA GCACAAAC AGCAACAAA ACNTTCTTA ATATCACTT GATAGCTGT 360

AAGACTGCG TTACTTGT GTCCAAATA TTTATGTGT ATATATATAT ATATATATAC 420

AACAACAC ACATATAT TCAACATA AGCAATAA TACATGCTT TCACTATTT 480

GTCTTTCCT GTTAGCAGA ACTGTATGAC AAGTTTAGGT GATCTTACA 540  
TATCTTAAT TCAATTAAT GTAAACAGA TTAAACAAA CAAAGAACT GTCTATTGTA 600  
5 GTGAGTCAT GCTTCTTAT ATATTAATCT GGTTCGGT TTCTCTTAAA TCTCACTTA 660  
TACTGTATC TGAATTTTAA TAAATAGAA TACTTACTT ATAAAAAAA AAA 713  
10  
(2) INFORMATION FOR SEQ ID NO: 73:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 73:  
20 GAAATCAGA GCTGTCCAT CCTCAGAC GTTTAGATT TGTCCCAAT TAGAACTGT 60  
GGACTATGT GTTCTGGCA ATCAGAGTC TGGAAATGG CTCTCAGGC TCTGTATGT 120  
25 GAGACGTGG TCACTTACC AGACATCAT CTGATTTTAA GCTCAGGCT ATTCACAT 180  
GCTGGCCAT GCTATGATT AACAAACAA AGCAAAATCT GCTTTTAG TTAGCAAC 240  
30 CTGATAGAA CAGTATTTT CAGATCTTT GGAATAGCA GTTCTGCAT TTTAAATGG 300  
GACTCAGAA GTGATGTCT ATAGTTCTCA AATACAAA ATGTATCTT TGCACAGAA 360  
AGAGAGCCG TGCCTGGAC TTGAAAGAT ACTGAGATC ATAACTCTTA TGAAGAAAT 420  
35 TAGGCTCTGT GATGTATAC TAAATATAG GTTAGAAG CATATGAC CTTTGTCAA 480  
ACTAGCTTC ACTAGAGGA CTTGTCTCA TAGAAGATA TGTCTTAAA GTATCAATTT 540  
40 TCCACAGTC ATGATGAGA AAGTTTCATT TCCACAGAA TGTGTATGT CACATAC 600  
AGCTGACAT ATATACAT AAGTTTCT GTAAACAGAA GTTCTTCTC TTCCATTTCA 660  
45 GGATCAGTC AGAGCTAAA TATTCATGT TTCACTTTAG AACTGATTC ATTTAGAAA 720  
GGAGATCTGG ATATTTTGG AGGTAGAAA TGAAGCAT TTCTGGCAT CTCTCTAAA 780  
AAGTCATAT ATATACATA AGTATAAA AGGTCTCTT TCACTCTTT TTTTCTTAG 840  
50 CATGCTTAC ATACTGTG CC 862

55 (2) INFORMATION FOR SEQ ID NO: 74:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 4602 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 74:

5 GCGAGGGGC GCGGAGACA GCGCGAGGC GCGCGCTCC GCTTCGGCG CTGAGACTA 60  
GCGGTGGGG GACGGACAG CCGCATATCC GCGGGAATAG GAGAGGCGA GACCGCGGA 120  
GCACAGCAC CAGAGAGGG GAGAGCGGC CAGCGCGGT CCGAGGACC GCGACACAG 180  
GCGCCCTGAG GCGGTGGCG GCGCGCTCC AGGACCACT AGACCGCG TCGTAGGCG 240  
AGGTGGGAC CAGCGCGCG GCGCGTCC CCGCGCGCG CAGCTGTTC CCGCGGAGA 300  
15 AGCGCGATA TCCCGCGCG CAGAGGAG CCGCAGGCG AGCGCTCCG AGCTTCGGG 360  
GCACAGCCC AGCGCGGAA GTCCCGTCT CAGTTCTAG GCGAGAGAG TCCCGGACTC 420  
20 CTATCTTAG AAAAGTTAC AACTATAA GATCCAGG AGAAAAA GCGAGAGAC 480  
GATTCGCC TCCCTCAGA AGTTTCAT GTTCATCTA GACTAGCG GCGCTGGCT 540  
AGTATAGA CATCTTTTC TCGCATCTT GATACAGA ACACCGAG CTCTCGCTC 600  
25 AGACCGTCT CATTCAGCT CATTCGAG TCGAGGAA ATACAGCA ACTGATAT 660  
GATTTGAG AAGACATCA GTCTCAGT GCGATTATA GTAGAGGA AGAGAGAG 720  
GAGAGAGA TTTTATCAG TGAAGGAG ATACATCTA AGATATCC AAGATATAG 780  
30 ACCTAAAC CCGACTTAG AAGGGAAC CCAAGGAC GAGAAAA TCAGAGGATA 840  
AAGAGAGA AGGAGAGA GGAATTTAA GTGAGTAG AGTTGAGT GAAAGAGAG 900  
35 GAGATGAAA TTAGAGAGA TGAAGACT CAGAGAGA GAGAGAG AGCAAAAT 960  
GACAAATC CAGTTTACC CAAAGAGA AAAAGGCT CATTCCATA TGTCCCTGT 1020  
40 GAGATGAG GATTTGAG TGTCTTCC CATCTGCT ATTTGCACA CCAATTTAA 1080  
TACCGCAT TGTCTAGAA GAAATTTTA TGTCCGATC CTTCTGTGG AGACTCTTC 1140  
45 AGCTTCTGA AGCACTTCT GCGCATGCC AAACATATA CAGTCAAG GATTTATTC 1200  
TGTGATATT GTCTCGGC CTTCAGAT TCCGCAATC TGGCATGCA CCGATTAAT 1260  
CAGCTGGG AGAGGATTA CATTTGAGA TCTGTGATT TACTTTGCA CAAAGGAT 1320  
CTCTTAAT GCAATGAG AAACATGTC CAGACTCTT CTACCAAT TCTTGCATA 1380  
50 TCTGTGGA AAAATTTAG AAGAGACA GGTATGCG ACACAGGA AAAAGGACC 1440  
CTGAGTCT CATTTGAGA GCTCTGCT CCAATGAG GCGCTCAT ACAGAGAG 1500  
55 ATATCTGG CACTAACCA GATCTCTCA CCGAGCTTC AGATGTCAG GCTCTCTC 1560  
TTCTCTGA GCGCTTGG AACTCACT CTGAGATG CTTACTGTTA GAGCTGAG 1620  
GATTTTAAA GTCACTATC AGTGGAGC AACGGTAG CTTGATGCT GATGGAGA 1680

1740 TCTTTGTGG AGCGCGACG AGTGGAGCA CTGAGGECT GGTATGAC TCAGATATAC  
1800 TCGGTCTAC CACAGAGGTT CTGATTAAG ATTGAGATC TCGCGAGCTT TATGGACAG  
1860 GAGAGTGG GGCATGGAC AGCTGAGCT TTGTATTAA AGTTAATAA GGCATATAA  
1920 AAAATCTAA GCAATTAAAA TCTATGAAA TAACTGAGG GCTGCTCTT TCCATTGCG  
1980 ATCAGACAC ACACATACAT ACAGCTCCA CTTCCCATC CCGTGTCTC CTTCTGTTC  
2040 TCCCTTATA AAATTGATG TGTCTTACC AGAAGGTAG ACATAAAGA AGCGACGA  
2100 GCTCTTAAG TGAAGGTAT TCTGATCTC GGTTCAGCC ATCAGAGAC TTCTGCTCA  
2160 TCGGCAATC CCGTTTCCA ACCTTAACT CTGATGTCT CTGATACG TTTTAACTT  
2220 TAACTATA TTAATCTCT CTAAATGCT TCTCTCTC TACTGTGC CTATGTTCT  
2280 GCGCTTACC CCGTGGGCA CACTTATCT CAATATCCAT AGAATCTAA TCTCTGAAT  
2340 CATACTCTC CAGTGGCTT TAAAGAAAG TGGTCTCTAG CACTAACAA ATCCTACAA  
2400 TACCTTAGT CTTTTTTGA ACCCTTTTA GCGAGNATG TTAGTTTCAI GGTAACTAT  
2460 ATGCTCTTG AGTTTATAC AGTGTGAAA CTTAAGATTT TTGAGAGGTT GAGAGGTT  
2520 GTTCAGATC TAAATACAG ATGATGATG GTTCTCTTG AATTGTCTT TTTTCTTTT  
2580 TTTTGTGCC TACATTTCC TTACTTTCC CTGCGGGCC ATCTGTGCT CTTGTCTTT  
2640 TGTCTCTCC TTTGCTTAT CAGTCTATC CAGCTCTCC TTAGTGAAG ACACCTCTT  
2700 TATGAGGA ACANATCTA TGAATCTTA ATTTTATG TAAAGAAAC TGTCTCTTT  
2760 CCGCTTATG AACTCTTG AMAGGAAA ACTTCAATG CCAAGTTAA TAATCTTAA  
2820 TAAATATCC TTTGCTCTT ACCTAATTT CTGGCATCA CATTTCTT GCGATGAG  
2880 TTGTTTGG GATGATG CTATATCTG TTAATGGAG AGAAGGTA GTGTTTCT  
2940 ACAAGTCAI ACCGCAAG GCGCAATC CTATTTGCG TCACTCTCAG GTAAAGATTA  
3000 ATTCTATCC TGTGTGCTC AGAGCTAGA ATCGAAGGT TACCTATTC ATTGTATT  
3060 GTCCAGATG CATGATGCT CTGGAAGA ATGAGTTTT GCTGMAAA AAAAAAABA  
3120 CAGTTTGTG TTCAAMACA TGGCTATCA ATTTTCTCA AATATCTTT TTTCCAAA  
3180 AGAGGATAA CAANATCTA TTTTGAAG AGGCTTACT TATACAACT AGTGTACGA  
3240 TTTGGATCC CAGGAAACAG AGAGTGAGC ACCATCAATC ACCAGTCTCA AATGCGTAT  
3300 TTTTCTTTT CAGATGTG CAGATTGCC ATTCTGCTT AATATGGGA TGAATATG  
3360 AATAAGATA GAGGGATGT AGAATATCT TTCTGCCA CATGTTTGG AGTCACTTT  
3420 GGTATATGA CTAGATTGA AATACAGA TTGATGAT GATCTACAA AATGTTTCT  
3480 CTCTCTCAG GTCCCTTTA CACTTTTGA CTAACTAGA TCTAATTC ACACTTAGT

3540 TTTTCTAC ACTATGCTT TGTCTCGTA AATTCAATTT GCAATGTTA GTCAATCAAT  
3600 AATTAGCA CTAACAAA AGCAACTGC ATTATTAAA ATCTTCTCA CATCGGAA  
3660 AATGATCT CTTTCTGAT ACCTCTCTC CACAAAAA ACACTATGT AGTCTACTA  
3720 AATGAAACT TGTGTACTT TTTCTTTCT TTTAGGGTC AGGACCTC TTTATAGTA  
3780 CAATTTGCT AATTAATTT ATTGACGAG TTTCGAATAC TAAATATTT TTTATAGAT  
3840 TTAATTTTT CTTTGTGTA AAGGATGCT GCAATGAGA GTTGTGTAA TTAACATATC  
3900 TGAAGGTTT CCGTCTTTC CTTCTGCTC CATATGCTC ATTGTCTTC CAGGAGCTC  
3960 TTTTAACTT AAGTCTTAC ATTCAATCT CTATGCAAA TTCTGTACC TTTTAAATTA  
4020 CTTCTCCAC TCAATATTC CATCTGAAT TGGTGTCTT AATTTCTGA ACTGTATGT  
4080 AATAGACT AATTAAATTT TCTGGAGAT GTGCATGCT CTTCTTTCT GTTGGCAG  
4140 GTTGTTTTC GTAACTAGA CTTCTGATA TCTTCAGAG AATTAGGCA AACACTGCC  
4200 ATGCGCTGG CACTACTGG AGTAATAAA AATATGAG ATATGACTA GCATCCAT  
4260 AGAGCACTG AACTCTCTT GTACTGTTT GCGAANAAG TATATGAT GTACTACAA  
4320 TCTACTAG ATATATATG TCTGTGTTT TCAATACCA TTTTCTCTC CTTTGTGTT  
4380 TTTCCACTT TCAATATAC TCAAGAAAT TCAAAATG TAAATGATCA AATTAAATA  
4440 TTTTATTTT TAAAGGCTT TTTGCTCTT TGTATGCTC AGGACCTTC TCTTTCTG  
4500 GAGAGAGG GTATGACTT GATATAGGT TGAAGGTT ATGTAAAG AATTAATAT  
4560 AAAAAATA CTTTCTTTT CAATCTTGT TTTTCTTAA TTCTAGTAA GCAATATTA  
4602 AATTAATAT GTAAAGACA AATTAAGAG TGTCTCTAT GG

## (2) INFORMATION FOR SEQ ID NO: 75:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1255 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 75:

60 GCGCGCGCGT TCTCTACA AATTAACAGA ACCGCACTG CCAAGGAGAG  
120 GCTTCCACT TTCAAATGAG TCCCTTGGG GAGCTAGCC TCACTGAT GATCTGCA  
180 AGCGCACTT TTAATTTTA TTTATTTTT ATTTTTTTT TAGATCTCT TTGGGGCTTC  
240 ACTTCAGAG CAGTTTITA AGGACACCA GAGCGGAGC CTGCTCTGAT TCTATGCTT  
300 GCTTCTACT ATAGAGTAA TTGCTTACT TCAATTTCA TCTTTTAA CAACTGTGT

5      GCCAAGAGT ATTGACCTT TTCAMATT CAGATCTGC CTCCTCGGAT AAMATTTTC      360  
CAGCAATGAG TAACTCTCTT CACACTCTG AAGCTCCGA CAGAGTTT TGCACATTC      420  
TTAGCACTGA ACTCTCTCTT GATCTAGAT GATCTCTTCC CCTCTCGAT GACATCTTC      480  
TGATGATCA GCTCCACAG AGCTACTTT GAGCGGACA ATCAGATCCA AAGCTCTTG      540  
GCTGTTTAT TAAATATCTA GTGTCACTT CTGAGTACC GCGCTTTAC AGCTGATGC      600  
CAGGCTGTG TGTCTTTAG ACCAGCTGC TTGCTCAGAG CCAATTTCC ATTTCATCA      660  
TTACTGCTT CACTGTGATA GTCACTCTTT TGAATCTGG GACCAAAAT GTGTATGTA      720  
TATGACTTT ATGTATGCC AGATTCATC CCAACCTTA GTCTTCGAA TGTATATTT      780  
TGATATCTT AAAAAATGA TTCTATCAT TACAGATTC AAMATTTGA AAGGATGTG      840  
TGTCTTTCT CCGAGCTCC CCTCTTCCC TTCAATGAA ACCACAGCG TGCATCTCT      900  
TGTATATGA GCGTATGCA CTCGAGGCA GTGTGTATG CACTCCCCC TTGTGTTAC      960  
ACAAGCTGT GGTGTTTAC CATGCTGCT TTTTCACTT AATATACAG CTTCGAGGA      1020  
TTTGTATTC ACHTATATA TCCACTGCG TCTTTTGTAT GGCACATTA TAACTACTGC      1080  
ATAATATGA TACGCTTAT TTGATTTAC TACTTCCCTA ATGATGACT TTTAATGTT      1140  
TTCTTTTTT TTTCTTTTT GTACTGCAA AGATGCTAT AATTAATGC CTATATCAA      1200  
AAAAAAAAA AAAAAAAAC NAGGGGGGG CCCCCGGAC NCAT      1255

35

## (2) INFORMATION FOR SEQ ID NO: 76:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 475 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 76:

50      GCGACGAG AATGTTTGA TTCTCTTTC TATTTAAGG GATCTCTCT CTCTGTATG      60  
TTGAAACTT ACTTAGTGA AGATGTGTTT CACATGCTG TTCTCTTTA CCTCATPAT      120  
CAGAGTATG CATCTATCA AAGTATGAT CTGTGGGATA GTTTTATGA GTTCACAAAC      180  
CAAACTAG ATGTAAAGG AATGATGAA ACTCGACCC TCGAAGAGG ATTTCTTTA      240  
GTGACTGTC AAAAAAGG AAGGAAGTT TTTATACAC AAGAGATTT CTTTATTAAT      300  
ATGAGCTGT AATTCGCC TTGATATCA AGGTATATC CTTCTTCTT TTATGCAAT      360  
CTCTTTTGA CTCTAGGTG GAATATTTT GAAGTGTTT ATATCATTA GTTCTGTGA      420

60

5      AACCTACGA GATTATCCT TCTTAGAAT ACTTACCTT CTTACCAAT TAAAA      475  
(2) INFORMATION FOR SEQ ID NO: 77:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 465 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 77:  
10      TTCTCTTGC TCTTCACTG CACCGACTC GCGGTGACC CTGACTCCC CTAGTCACT      60  
CAGCGTCTT GGCATGCGT GCGCGCGCG CGAACCTGG TCGCGCTTC CCGCGTGTG      120  
20      GCTCTGCGT TGTCTGCGT GCGCTGTGC GTCCCGCGG CCGCGCGCG GCTCTCTAG      180  
TGTCTCTCG CCGTGTGAAA CATCGATAC GTGAGCGGC AGACCAACT GACGCTGTG      240  
25      ACCCTCTCG AGATGCGCG CTTCGCGAC AGCTGCGCA AGAGCGCGC GCAATGCGCT      300  
GTGCGCGCT CCGTGGCGC CCGCGAGAG CTCAAGGCT KCGCGCGCA CAGCGCTTC      360  
TTCTGTGCG AGCGCGCGG CCGAGCGGC GCGCGTGG TCGCGTGTG GGTCTGTGG      420  
30      GCTCGACTT TCAAGGAAA AGTGTGTGT GCGCGCGCA ANGAA      465

## (2) INFORMATION FOR SEQ ID NO: 78:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1907 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 78:

45      ACNCGCGC CACTACGA TTCTTATGA ATTCTGAG GTTCCAGA GAATTAAGA      60  
AGACAACTG GACAGATCC AGAGAGCTT AGTGTTTTG GAGAGATA TTAGAGAT      120  
50      GAGAGATG AGTGTCTAT ACTCTCTCT CAGTGAAT AGCAAGTGC CTCATTTGA      180  
AGCTCTCTT CCAACACA GTAGTATAT TATTTACA GAATACAGC AACCTCAG      240  
TTCTGTGCG AGTTCTGGA CAAGAAACA GCTTGTGAT AATAGCACT TACGCTAAG      300  
55      AGAAAAACA CTACTGTC ATTGTGAGA CTTCAGCAG TTTTACTTT CTACAGAT      360  
GTCTGTATC TCGATGACA GTCGACTGC AAGCGATTC GATGATTTT AGGATGCTT      420  
60      GTCCAGTTT ACTGATATA ATTCAAGAG ACCTTAGGC ACNTGTGAT ATGCTATGA      480

540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560  
1620  
1680  
1740  
1800  
1860  
1907

5  
10  
15  
20  
25  
30  
35  
40  
45  
50

TCTCTATAT GATTGAGTA TACTCTCTAG TATTGAAAT GACCGGATTT GTGACTATTT  
TCCGATTTCT GCACTTACAA AGAAGTATA AGTCTATGAA TATGACACTG TCATCCGAGA  
5 TCAAGTGGAT ATTCAATACC CTGAGATGCA AATGACTCTC AATTGAAA TCAAGCTATAT  
CAGTTGGAGT AGTTACATA AGAAGCTTT AGTAGAGAT GATTATGAG GCACTATATAT  
10 TTATGGGAT GAAATGAG GACGAGATC AAGGATCTAT CAGAGGATG AGAAGAGTG  
TTGAGATGTT GACTTATAT TGAATGATC TAAAGCTCTG GCTTCAAGTT CTGATGATC  
15 AAGATGAG CTGTGATCA CAATCTAGA CAATCAATG GCAAGCTATG AGGAAAGGC  
TAAATGATG TGTATATAT TCAAGCTCTC TTCCAGATAC CAATTTGCTT TCGGCTGATC  
20 AGATCACTGT GTGACTATCT ATGATCTCTG TAACTATGAA CAGCAATCA TGTATATGAA  
AGGACAGCTT AAGCAATCT CTATGCGAAA GTTTGTGAT GTTGAGGAAA TTGTCTCTGC  
25 CTCAGAGATC AATGAGCTTA AACTGTGAAA TGTAGGAAA CCAATCTGCT TACGTTTCTT  
CAGGCTCAT ATCAATGAAA AAGCTTTTGT AGGCTGCTCT TCAATGAG ATTAATAGC  
30 AATGATATTT GTTATGCTGT CTCTCACT GTACTATGAA GCACTTTCTA AGACTTTCT  
GATTTGCTCT AAGATCTGAG GTACTATTA GTCTCTAGAA TTGATATGAA GGTATATCTC  
35 AAGTCAATTT GTACTGATC CTCTGATAT AATCTGATG GATGATGAG GAGAGAGATC  
AGAAATATCT ATGTGATCT TCTGCGAAA GTCAATCTG GTTTTGGAT TTATTGAT  
40 ATTTTCTCT TTATTCTTT TCTGCTCTT ATGACTTTTG GCAATTTGG AATACCGAGC  
CAACTCTCA CAATCATAT AATCTGATG AATATGCTCT TCTTGGTGT GTTATCTAT  
45 ATTGAGCCAC AGTTGAGCTT GAAAGTTTT TGTCAATGCT NCGAAGAT AACTCTTTT  
ATGAGATGAC ATATGATGAC TATATGATA CAGTATGATA TTTGTAAAT TGTATTTCT  
CAACTCTTT GGAATATCA CAATCAATA TAAACAAAT ATTAAT

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1168 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 79:

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1168

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

GTGGGGTGT CCCKCSGCC ACGATCTCA TCGTACTTT CATGAGCAC ACTCGGATCA  
CCCATGACTG ATGCTTATTA ATTGTGAAA GCGAAGGAC CATTATCTC CCGAAGCTT  
AATTCATGG GCGATTTCT AGATTTGAG GAGAGCTTA ACGGGTGT GACACCGAGA  
ATCTTACAC CAAAGCTGAT GCGGTGGAG AGGTTTGTGT GACATGCTC TCGATGAAA  
10 GAAATGCTGC TCTCAATAG GAGCAATGA GGAAGGAGA TCGATTTCTG TTTTCTTCT  
TTCTTTTTTT TTTTGTAGTT GCGATAGT TTGTGATGG AAGCAACTT GTTTAAGCAC  
15 TTATTTTATA ACGATGATA GAGACTATA ACTTTGATG CAATGAGAT TCACTCTCCA  
CAAACTGACA AATTAGGAG GTTAAGGAG TAATTTTTTT AAGCCACAA TAAATATATA  
ATCAACTTGT TTCTCTCCC TTCTCTTTT AAGTATTTG TAGATTTTAT GACTAATAG  
20 TCTGTGAGG TTCAAGACC GAGATATA CAAGCTTAA ACCAATTAA AAGAACAAA  
AGTAAATAGA AAGAGATG ATTCAGAG GCTGTGGATC AACTGGGCT GTCCACAGAG  
AAGAACAAA CCAACAAA CCAAGCCCTG TTGTCTCAC TGTGTCAAG AGAATATGAG  
25 GCAAGCTTAA GTGTCTTAG ATCTCTGAG GCAATCTTA AGAAGAAA GCAATCTTT  
GATTTGTGT GTTTCACTCT CTGATTTTT TTTTCTTCTC TTCTCTGGG TTTAAGGAT  
30 TTTTGTGAA ATAGTAGGA ACTGACCAT ATATGCTTTC ACTGCTTCT TGTGATTA  
TATGATGTTT TAGTGTGCA AAGAGTTAG AGCTGGCAGC TGAATGATAG ACAAATAGT  
35 CAATTTGCC ACTGTGAGA TAGAAGGAA TTCAACATA TATCAATATC TTTCTTTCC  
ACCTTTTCC TTTTCTTTT TTTTCTCTG TTTGATCTG GTTACAGTC CATTAACCTT  
GTTACATAG TATATGAAA TGTAGAAA AAAAATTTAT TTAATATAT TTTTGGAAA  
40 AAAAAAAAA AAAAACTGA GGGGGGCC

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1285 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 80:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 80:

60  
120  
180

55  
60

AGAAATCAC ATCTAAGCA AAGATCTGT CTAGACAGT ACATCTCTG TTGAATTC  
ATCTTCCAC AGAATTTCT GTTTTAGGG ATGAGACTAT TCTCTGCTTC ATCAGGAAA  
GAGAAATGTT CAGGTTGTA GCGATGCCAC ACTTATAGT TCTGCTCTGC TGAAGGTTTC

CTCCAGGACA GTTTCGTGAC AGCTCCGAAAT CTATAGTCCAT GGTCTAAATCC TTGAGATATCT 240  
CTTCTTTTCC TTTCCTGTCT CAGGATACAG CTGAGAAATTC ATTGGAATGAT CATGCCCTTA 300  
GCCCTTTACT GTGATTTGAT GGTTCGACTT TCAATTTGCTT TAGTCTTAGA ATCACTGTTT 360  
GACTCTCAG ACTTCACTTA ACTTTGGAAA CTCTCTTTTG GAGGCTTCTC ATTTCCTCCCT 420  
AATCTCTCC TCCCTGACC CTAGAAATTT CCACCAAGC AATTAATCCA GTTAGATCTT 480  
AGTTCTCCG ATCTAGTTGA TATTTAAACA ATATCTAGTT GATTAATTTCT ATTCAAGTTGG 540  
ATCCAGAAAC CAGTATCTCT TAAAGAACAC CTCTCATACC TTCTGACCT AATTTTGTGT 600  
GCTGTGTGT GTCCGCGCAT ATGTATATAG ACAGCGAAT CTTTTTTACT TTTGTAAAG 660  
CTTATGCCCT TTTCGTATCT ATATCTGTGA AGTTTATAT GATCTGCCAT AATCTCTGG 720  
GGACCTTTGT CTTCCTGTGA AATGTACTA GAGAAACAC CTATATTTATG AGTCATCTA 780  
GTTCGTTTTA TTCCACATGA AGGAATTTTC CAGTATACA CACTAACAA CTCTCCCTTG 840  
ACTAGGGGGA CAAGAAAG CAATCTGAC CATTAACAC AATTTACCTGG TGAGAGTTG 900  
CATTAACAGA ATTAGTATG ATATCTGAGA CAGATCATT AATCAATTTT GTTGTCTCC 960  
TTCCAAAAA CATGTACTGA CTTCCTCTTG AGTAATCCA AGTTGTTTTT TTTATTAATA 1020  
AACTTGCTCT TCATATGAT TTTCAAAGTG GTTGTGTGG CCATTAATTT GAATGATGG 1080  
AACTGACTGA TAAAGCTGA CAATATGA CTGTGCTTAA CAGCAACAC AGTAATTTTG 1140  
ACATGCTTAA TTCAAAATG CTATTTCTT TATTAATTTT TTTCCTAATA TACATTTGA 1200  
AACTATTTT CTGTATCCA AGAGCTGAGA TCTTAGATT TATGTAGTAT TAAGTGAAA 1260  
AATACGAAA TATTAACAT TGAG 1285

40

(2) INFORMATION FOR SEQ ID NO: 81:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

45

50 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 81:

TTCTCAAGCC CAATTTCTAC GGCACCGGA AGACGGAGGT CCTCTTTCTT TGCCTAACGC 60  
AGCATTGCTT CTGTCTCCA AGAAGCTCT GAGCGGGTG GCAGCTCCAA AGCATGTGAT 120  
GCTGTGAAA TTGACCGGTG TTTTCTCC TCCTCATCC ACCGTTCCCT ACATGTTGAG 180  
AGATGTCTC CCCCTATCA TTTTCTCAG GACAGACTT AATTAATCCC TGACAGGAGA 240  
TGAGTAAAG AAGATTGCA TCCAGCGTT CATTAATAATC GATGCCAAG TCCGAATCA 300

60

TATTAACCTAC CCTGCTGAT TCATGGATGT CATACGATT GACAGAGCG GAGAGAAATTT 360  
CGTCTGATC TATGACACA AGGTTGCTT TGTCTGATAT GTTATATAC CTGAGAGGC 420  
CAATGACAG TTGTGCAAG TGAAGAGAT CTTTGTGGC ACAAAAGGAA TCCCTCATCT 480  
GTTATCTAT GATGCCGGA CATTCGCTA CCCGANTCC CTCTCAAGG TGAATATATC 540  
CATTCGATT GATTTAGAGA CTGCAAGAT TACTGATTTT ATCGAATTTT ATTCAACCG 600  
CGAGGTGTC TGTTCACCT AGAGCTTGG CAGACTCTG CCCAGGCGAG GACTGAGGCA 660  
AGCTCTAAG CACTTCTAGG AGTGTCTCT TCTCAACAG ATGAACTCAC TGTTTTCTTG 720  
GCACTACTG CTTTTCTCT GTGCCACCA CTTTGGGAG CATTAGAAA AGTGTGCTC 780  
TGTGGGAT TCTAGACCA CAGGCCAGA CTTAGAAATC CTGGGCTTC TGGCCGCGG 840  
GGAGAGAGC CTGCTGTGA CCGAGAGAA GCGACTGCT ACTGCCAGG TGAACCTTGG 900  
GGGAGCTCG CTGTCCCGC CCCGCGAGG CTCCGAGAG CCCGAGAGC CGGCGCTGTC 960  
GGCCCCCAC AGCTGCCAGA TCCCGGACC CAGAGGCGG GTGCTGTTC AGCGGAGAA 1020  
GCACTGCGC AACTACACT GGAATCTCTT CGGCTGCGC TTGCGCAGC GCGAGCGGC 1080  
ACAGGGAAC CAGCGCGAA GCGTGGCGG GCGCTGAGG GCGAGGTTC GCGCAATGAA 1140  
CTTCAGACC CAAGAGTTC AGAGCATCG GGGCGGGGC GGGCGCGGG GACTGAGGC 1200  
TAAAGGAGG GCGCTGAG CTTCACCC GAGCGAATTA AAGAAATTTT GGTAACTCA 1260  
AAAAAAAAA TCGGGGGGG 1290

(2) INFORMATION FOR SEQ ID NO: 82:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 684 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 82:

TTTATTTGAT TCTGTAACTA TGAAGTCTT AATTAAATCT TTTTGGACT TGTAAATTTG 60  
TCTTTATGG TTTTAAATCT CAGCTCGAG TTTTCAATAT TACTTATCC CTTTGAACAT 120  
GAATTTTAT ATGACTCTT ATGATTCAAA AATCTATCAT CTTTGGTAG TCTCTTTTAT 180  
TTTTCACATG TTTCCTTTGA TTTTAACTCA TGTATTTTAA ATCTCTGTT WTTTTTTTC 240  
TTTAAAGA CATCTTTTA AAAATTAATT GAGGAATAT TTGATTTTAA TGAACAAAGC 300  
ATTACTCAC AGAGAGATT TTTTGTGTTT ACAGTGGC TARGAATCT ACAATGTCG 360



420 GAGCAGTAG AACACAGGT GATGAGACA TCTTGGGAT CTTGTTTAC TTTCGGCAT  
480 CTTTCTCCC AACCTCTGG GAATATCAT YCATATCTTA RTCTAGGCT ABAAGTGCT  
540 TTATCAGAGC CCACTTCCA GGGCTCTGG CTTTACTAC TGTACCCCA TCAATACTGA  
600 GCTTCATGA TTGATCTCT TTTTACTTT CAGATTTCT TTTAAMATC TTTTATTTT  
660 TTTTCTTCC GAAGATTC CCAACATTA CCAATCCCA CTTTGGTTC AATTTTTC  
684 GCTCTCATTT TGATTTTTC AAGA

10

15

20

25

30

35

40

45

50

55

60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2024 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 83:  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 83:  
CTCCAGCAAT TGGCAGAG TGGCTGGAG GCTTCATCTT TGGCGCCT GCGTCCCT  
TCTTGGGAT GGAATCTGA GCTTCTCTG TTGCTTCGTC GCGGGTTTC GCGCTTCTC  
GCGCTCTGG GCTTGGAGC TGGGAGGG GTTGGAGGG GCTTGTATC GCGCTTTTA  
AGTTGCTCT GGGGGGCGA TGTGGGCGG CGAGATCGAG GCTCTAGTGT GCGAGCTGAG  
CGCGGGAGC GAGGGGATG AGGAGGAGA GTGGCTCTAT GCGATGAAA ATGAGTTTGA  
AGGCGAGAA GAGAAATG CCAATGCTAA TCTCTCATCT GGAATTCAG ATGAAACTGC  
TGAAATTCCT GTACAAAC CGAAATGAC TGAGCTGAA GATGATGAG ATGATGACAG  
CGATATGAT GAGATGATG TTTATGTCAC TATAGGAGAC ATTAAACGG GAGCACACA  
GTATGGAGT TATGTTACAG CACTGTAAA TCTTACATC AAGCAGGGG GAGAGTTTA  
TGAACTACA GGGCAAAAG TCAAGAGT AGACTTCAT GCACTCGAA GATTAATGG  
AGTTCCACT TTAGAGTAG ATTGGATTC TTTTGAAT ATACCTGCG GTAACTGCG  
TCTGATCTT TCTGATTTT TTAAATATG GTTTATGAA GATTACTGGA AGCTTACTG  
TGAAAAACAA AAGAGTATC GATGGGACT TGAAATTTA CAGTAACT CTACTACAA  
TAAATTTAG GTACGGAGG GAAGACTGG AAATCGAG AAGAAACTG CCGTTCTCT  
TACAAAGCT GAGTTTACT CTTCTCTTC TTGTTTCAG ACTGGGCTTC CACCGAGAG  
GAGTTTACT GGGCAATG ATGTTATCG TGAGACTTAA ACTATCAGCC GATTAAGAG  
CAGGCGGG GCAATGAGA ACGACATAT ACGATTCCT TCTGAAGAT GTCCTACTGA  
AGTATGACAT AATTTTACG AACACTTC GTTTTTCCT CAGGAGCTC CTCACACTCA

1140 CTTTCACCT CTTCCATTTT TTCCACTCC TGGAGTGTG AGCACTGCTC CACCTCTGAT  
1200 TCCACAGG GCTTTTCTTC CTCACAGG CCGTCCACT CCACTCTCTTA TACACATAT  
1260 AAGAGTGA CATCTCTCTG GTTATGATAG TGTGTTCTGA COTCATTTT CATATGCCA  
1320 TTTTCTTT CCGATCTTC CTGGTCTGC TCTTGTGTG CTTAGTCTTG TGGACACAG  
1380 CAGCATGG GACTATATG CAGAGACA GAAGACCA GATAGAGAG GAGACAGAGA  
1440 CAGAGGGA GACCTGATC GGGAGAGA AGAGAACGC ACCAGAGGA GAGAGAGGA  
1500 GCGTATAC ATGCTTACAC CAGTGTTTT CAGACGAT GAGAACCAT ACAGATNAG  
1560 GGAATATGA GAAGAGCTT ATGACCTCA CAGAGCAT CAGAACCAT CAGAACCAT  
1620 TAGAGAGA CAGCAGGG AGAAGAGCA AACAGCAT AATCTCTCTC GAGTATNAG  
1680 TAGAGTGG CATGAAATG AGAGAGGA TACTCAGAG AGACACAGC AAAAAATC  
1740 TAAAGAGC AAGAGAGAA AAGAGCGG CAGTGGGCT CCGCTCGAC AGAGAGCAG  
1800 CGAACTACA CTTCCAGAT AGGCTATGTT TGGCTTTT GTGTATNTTA GTACAGAG  
1860 TAGATACTAT AATCTCTTT ATTTTCTGG ATATGTTTA AAAAAATTA CTTAAATCTT  
1920 GTCTGTTTG TTAGTATGAA AGTTTACTT TTTTTCAAA ATAAAGAGT GATTTTTC  
1980 TGTATATTA AAAATCTTG TCTTACTA TTTCAAAAT AAAAGAGAG CATGACTTT  
2024 ATATCAAAA AAAAAAANA AAAAAANA AAAAAAGGC GGC

5

10

15

20

25

30

35

40

45

50

55

60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 931 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 84:  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 84:  
CGCGCATTA CCGGAGCGG GATCTGAGCT GCGAGATGA ATGTGGGCT GCGACAGAG  
GAGTAAACC CAGCAGCGG AGTATGAT ACCCGAGGA TCTGGCTGC CTACATATC  
TTGGTAGAT TCGTCAATAT GGTCTACTC AGCATCCCT TCTTAGCAT TCTGTGTC  
TGAACTTGA CCACTGATAT CATTACTG GCTATGATG TCTTCTTTC TACGATGAA  
GCGACACTT TTGAGACTTC TGACAGGA AAGCTGGC TACTGACACA CTGGAGCAA  
ATGACTATG GCGTCAATTT TACTCTTCC GCGAATTC TACATCTC TCTATTTTG  
CTCTATCTC TGGCAGCTT CTATACAG TATGATGCTG CCACTTCTT CATCAGACA

5 GGCCTCATTC TAACTGTACT GCTCCGGAAG TTCCCCCACT TCCATGGGGT TCGTCTCTTT 480  
GCCATCAACA AATACTGAGG GATGGTGTIT GGGACAGCTC CHTGGGCATG GGGAAAGCAC 540  
TGAACAGAG GACTATAAA CATCTCTCTC TTATTTCTCA TACTGTCTTC TACACTTTTA 600  
AAGCTTGGA ACTATACAC CTTTCCGAGA CTTCCGAGAA GAGAGAGAT TGGCAATATG 660  
GGCTCTGGG CCCAGTCTG CTAGTGCCAA GTTCTTTTGA ATCAGGAAG CAGTGAGGT 720  
AAGGCGGAAA TCACTCTCTT CATTACAGG AAGCAATTTG GGCAGCTCTT TTGTCATTTA 780  
CATCTTTCCA TATCTTTTAC ACTTACACCC TTCCAGCTCT GTTTTGGCTGT GTATTTTCT 840  
TCAATTAAT TTTTTCAGCT ATAGCTCCAG TTATATCAGG ATGGGTATAG AGCTGTCTTC 900  
ATAAGGCTGG GGTGGGAAG ATGGAACTACT G 931

## (2) INFORMATION FOR SEQ ID NO: 85:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 825 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 85:

30 CCGGCGCCGC GGGGTCTTCA GGTACCGGG CTGGTTACAG CAGCTCTACC CTTCAAGAG 60  
CAACATGCG ACCGAGAG GACGAGAGA AGATGCGGA GCGGGAGGG CTGAGCGGA 120  
CGACCTGCT GCGGAGCTG ATTCCCTCCG GTCCAGGCCG GAGTGGCTG GAGCGCGCC 180  
GCGCGACCAT CCGGCGCTGG AGCACTTCCG TGGACAGCA GCGCTTCTCA CCGGCGCGCA 240  
ACCTGGGAGA GCTGTGGCAG GCGCTGTATC GCAAGGTGGA GTACTACAG AGCACTATG 300  
TGTTCGTGTT CTTGGGCTTC ATCTGTACT GTTGGTGAC GTCCCTATG TTGCTGGTGG 360  
CTCTGGCTGT CTTTGTGGC GCTGTATTACA TTCTATATCT GCGACCTTG GAGTCCAGC 420  
TTGTGCTGTT TGGCGGAGAG GTGAGCCCG CCGATCAGTA TGTCTTGGCT GAGGCGATCT 480  
CTTTCCCTTT CTTCTGGCTG GCTGTGGCG GCTGGGCGCT CTTCTGGTG CTGGAGGCA 540  
CCCTGTGGGT CATGGGTC CAGCTGCTT TCAACAGAT TGAAGCTGTG GAGCGGAGG 600  
AGCTCAGAT GGAACCTTG TGAAGTGTCT TCTGGAGCT GCGGCGCTCC CCGGCGAGCT 660  
GCCCAACCCC TGGCCATGCC TTCTCTGCAC GCTCTGTCTG CTGGGCGCCA CAGCGCGTTC 720  
CATTCAGAG CCCGCGAGG GATCCCGCTT TTGAATATTA AGCTTTATG GTTGTCTATC 780  
AGAAAAAAAA AAAAAAAAAA GGGGCGCTTC TAGGGGTCAA AGTTA 825

60

## (2) INFORMATION FOR SEQ ID NO: 87:

- (1) SEQUENCE CHARACTERISTICS:

## (2) INFORMATION FOR SEQ ID NO: 86:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1238 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 86:

5 CATTAAAG GATGAATAT GACTCTGCT GTTTTTTTAT TTCTATGAG GCACTTCTG 60  
GGAGCGTTT CTGGCTCTCA GGTCTGTAGA AGCTCAGTT TATGATGTC TCTGTGTGTG 120  
CTGCCACTA CTGGAGAGC CATAGCTCC AGCTTTAGA AAGCGAACC CCGGCGAGAG 180  
TGTGGGAG TGGATGCA GATGGCAGG GCTTTGGAA ATGAGAGTG AGATTTTTC 240  
CAGAGAGGT GTAGAGAGG GATGATCTT GATACATGA TTAGGATCT 300  
GTCTGGACA CTGGCTTCC TGTCTACCTG CTCTTCTCT CTCTCTGCT CCGAGAGAG 360  
GCTGGCTAC TGTCTGCTT TCMTTTCCA GAGCTGCTG CTGCACTAC ACTTATGCTA 420  
TCTTCTCTA CTTTCTCTT TTTCGCGAT AGTGGAGTG ACGAGATGT GATTTGGGCA 480  
GGATGTCT TTGATGCAT CAGACTTTA GCTTCTGTT GCTGTGTCT CAGCTCTGAT 540  
TTGATTTCA GCGGTATGG AATGTTTCCA TGGAGCTGA GACTCTACT GAGCTGAAA 600  
CCTCAAAATG AACACATCC CTGCTTCTCT GCGAAGGATC CTTTATGGGT ATCCCGAGCT 660  
TCCCACTCT TTTTCTGTGT CTGTACAGAG AACACAGAG TACTTGTAT GCGCTGTAC 720  
CTGGCAGTT GATTTTCCC TCGAGGCTTG AGCCGAGCC AGAGCTTTGA AAGGTATTC 780  
AGTTTGTTC CCAAAACT GAAAAAAT GCGTGGGCC TGAACCAAT AGCTTGACC 840  
CTGTAACT CATTACCTG ACCCTCTTGT TTGATATTA CCGAGGTAGA ACAACTCTCT 900  
CTCACTGCT GTTGTAGA TACGCTTAG CCACTCATTT AATGATATC TCTTATATA 960  
TCTTTTGGC TGAATACCT GCGATCTTT TTCTTGGGC AATCTATCT TTTTCTAGA 1020  
GTTTCCAG GCTTACTGA GGACTTAC ATACTCTTAA TGGCTTTCT CTCTCTGTT 1080  
TTACTTATG CCTTCACTC CTGATTTTAC CTTCCAAATA CAGGATTCAC CTGTACCCA 1140  
GCGCTTACT TCAAGATAC AGATCACT GTACCGAGC CTTTACTCA AGCTCTGTT 1200  
TGGAGAAC CAATATGA CATCTCTCT GTTGGCT 1238

- (A) LENGTH: 1460 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 87:

```
5  ATTGCTCTCT GATCTCTCTT GACACTGGGG TCATCTTTCA TCCCGGAGG GCAATTTCTG 60
10  CTGCTCTCTC TCACCGGGGG CTTGCTGGGG GTGCGGGAGG CCAATTATTC CACCATCGGG 120
    CCGACTCTCA TTCCGACT CTATTGTGGC GACACGGCG ACCGATCTCT CAGCATCTTC 180
15  TACTTTGGCA TTCCGCTGGG CAGTGTCTGG GCTACACTTG CAGGCTCCGA AGTGAAAGAT 240
    ATGGCTGGAG ACTGGACTG GGTCTCTGAG GTGACACGG GTCTAGAGAT GTGGCGCTTT 300
    CTGCTCTCTT TCTCTGTAT GTGGAGCGG CCGAGGGGAG CCGTGGAGCG CCACTGATAT 360
20  TTTCACACCC TGAACCCGAC CTCTGTGTGG GCAGATCTGA GGGCTCTGGC AAGAAATCTT 420
    AGTTTGTGTC TGTCTTCTCT GGGCTCTCTCT GCTGTGGCTT TTGTACAGGG CTGCTGTGGT 480
25  CTGTGGGCTC GGGCAATCTCT GCTGTGTGTC GCGTGTGTTG TTGGGGGAGC CCGACCTCTC 540
    CTGTCCCGAG ACTCTCTGCT TTCTCTCTGAC AGTCTCTATCT TTGGACTCAT CACTGTGCTG 600
    ACCTGAGTCT TGGGTGTGGG CTTGGGTGTG GAGATCAACC GCGGGCTTGG CCACTCAAC 660
30  CCGCGGGGCT ATCCCTCTGT CTGTGGCACT GGGCTCTGCG GCTCTGCAAC CTTCCTCTTC 720
    CTGTCTCTTG CTTGGGCGGG TGGTAGGATC GTGGCACTTT ATATTTTCTAT CTTCATTTGA 780
    GAGACCTCTC TGTCTATGAA CTGGGCAATC GTGGCGGACA TTCTGCTCTA CCGTGTGATC 840
35  CCTACCGGAC GCTTCACCGC CAGGGCTTTC CAGATCTGTC TGTCCCACTT GCTGGGTGAT 900
    GCTGGGAGCC CTTACTCTAT TGGCTCTGAT TCTGACCGCC TGGCGCGGAA CTGGCGCGCC 960
40  TCTCTCTCTT CCGATCTTGG GGTCTCTGAG TTCTCTGCTA TGTCTGTGCC GTTTGTGTGG 1020
    GCACTGCGGG GCGCACTTCC TGGGACCGCC CATCTCTATT GAGGCGGACC GCGCGCGGCG 1080
45  ACAGCTGACG GTGAGGGGCT TGTCTGACGA AGCAGGTTCC ACAGACGACC GGAATTTGAT 1140
    GCGCCAGCGG GCGCGCTCCA CCGCGTGGCC CCGTGGCAAT GTGCTCATCT GGAGAGGCTG 1200
    CCGCTCACTT ACTGTGCAAT CTGACAGAG TGGCGCTTGG CCGACCCGAC GAGGGGCTTG 1260
50  GGGCTTAAGC CTTTGGGCTG GCGCAGCTTC CAGAGGAGCC CTGGGCGCTG TGGCAGCTCC 1320
    CAGACACTAC ATGGGTAGCT CAGGGAGAGA GGTGGGGGTC CAGGAGGGGG ATCCCTCTCC 1380
55  AACAGGGGCA GCGCCAGGGG CTGCGTCTCA TTGTGTACGG GATTATTAAT TTGTGGCGGA 1440
    AAAAAAAAAA AAAAAAAAAA
```

60

(2) INFORMATION FOR SEQ ID NO: 88:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1395 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 88:

```
5  CAGGTGCGAA GTGGAAATGG TCAGTCTCTCA GTCTGTGGCT ATTCGGCCAC GTGCTGTGGG 60
10  GAGATGGGAC GCTTGAGGCT CAGCAGCGTG GAGTGTCTGGC CTTTGTGCTC CAGCGGTGGG 120
15  AAATTTGGCA TTTCAGGGCC GGGAACTGGG ACTCAGGCTG CCGCGCGGAC GTTTCCTCAT 180
    CTTCCACCGG ATCTGTGGGC GCTGGCACTG GCGTGTATGT AGTTTCTCTGA CTTCTGAGCC 240
    GTATTGTCTC CAGATTAAAG GTACGACATTT TGGAGGCGCC AGCGAGAAAC GTCACCGGCA 300
    GAAGCTTCAC CCGCGGAGAG CGGCGCGGCT GTGTGTGCTC CCGAGAGGAC AGCGAGCTTG 360
20  TAGGGGGGAG TCCCACTCTGA AAAAAAAAAA TCCAGGTCCC CAAGGGTGA CCGTCTTTCCG 420
    GAGACAGCGG ATCCGACTAC ATGTGGGTGC CCGAAGAAAT TTCACTCTTG AGTCTCTCAC 480
25  TGTCTACCCC GGGTCTCATTT CCGAGAGAA GCACTCTCTC CTCTTTGGA GAGACTCTAC 540
    ACCGTCTACA CAGTGGCAAC GGTCTCTGAG GTGATATGCA TTCTGTGCTG GATTCATCAC 600
30  TCCGCAATCA AAGAGGCCAA CAGAGGCCAA CTAGAAATAT GGGTCTCCGAG GGTCTGGTCA 660
    GCGCCCTTAA AACTGCACTT AGTTGGGCTG AAGCCNTTAG ATTATTTCTT TTTCTTATTT 720
35  TTGTAAACGA ATGCTATGCT TCTGTCAACT TATGTATCTT AAGACTCAAT ATAAACCCCT 780
    TGTATTAATC GAGGAAATCA ATGATTTGAT TCCCGAAGAA CAGAGTGGGG GATGTATATG 840
40  TCGAACTGCG TTTTATCTAA CCGTGTTTT AGACTTTTCC TTTCCTTTAA TCACCTAGCC 900
    TTGTTCAC CTGANTTAC TCTCCTTAG CTAGAGGCG CAGATGGACT CCGTCTGGC 960
    TCTTTCACT GCGAGCGGCT TCTTCAGGG ACTTAACTTG TCGAGCTCA CTCCGAGCAC 1020
45  ATCCAGATAT GCAATTAAT GATTAATAC TGTGGCAGC TATATCCGA GTTCCGAGA 1080
    ATTCGTCGA TTGATTAAC CCGAAAGCCC GCGCTCTATC ACCTTTATAT AATCTTAAGG 1140
50  CCGCTGAGC TGGAACTATT AAGCTTCTG TACCACTTTA TCTTTTAAAC TTTTGTGCT 1200
    ACTTTATTTT TGTAAATATG TTTTAACTAG ACCCGCGCTC TCGTTTCTAA ACCAAAGTAT 1260
    AAAGCAAT CTAGCGGCTT CTTCAGGCGG AGAGATTTTC GAGCGTTAGC CTTCTCTTGG 1320
55  CACAGGCTA AATTAAGCGA TTCTCTATGT GTAAAAAA AAAAAAAAAA CTCCGAGGGG 1380
    GCGCGCGGTA CCGA
```

60

## (2) INFORMATION FOR SEQ ID NO: 89:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1186 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 89:

5 GGCAGAGCC GGCAGAGCC GCTAGAGTCA AACTGCGGG GGCAGAGCA TGTNACAG 60  
AAGACAGAA GATGAGACT TGTTCATCA CTTTCTTAG GGCATCTTG TGTTCATCTT 120  
TCCCTCTCC ATCATACCT CTCCTCTCG GAGCTCTCG CGCTTGGCT GTANTGTGG 180  
CACTTACCTG GATTTTTCAG TGGAGGATG AAGGCGAGA CTCACCTTAC GCGGTGGAC 240  
AGATGGGAG AGGAAAAAG CAGAGATGG CAGGAGGG GTCCAGACA AACGAGAG 300  
GTTGGGTGAG GGGAAAAAG TGGGAGAAA GAGGGGTGCA GGGCTCTGAG GCGGTTCAC 360  
CAGCAGCTGC GGCCTCTGCG GGCCTCTGCG ATCCACTTC GAGAGAGGG TACCACTTC 420  
CTGGTGTGTA TCAAGAGTT TGTTCATCA GTGTATGCA TGAATCTCT AGGTTTACA 480  
AGCCTCTG GTCCAGCAT ACATCCAGC CATTATATA AGNTAATA TATATATA 540  
TGTATGAAA TTATGGGAG AGAATTTATA GCACTGAGGG CCGTCTGCG CTGCTGGAC 600  
AAGCAAACT AAGCTCTTGT GTTTGGGAT TATGTTCTG TTTGTTATT GTTTGTTT 660  
GTGGTCTG TTATGTGCG ATACGACAG TCGAGTGGG ATTGCTCTGT ATTACAGAT 720  
AGTGTTTTTA ATTCAATAT GTTCTAGTGA ATGTCTACT CAGCAGCTCC TCTTAGCTTA 780  
ATTTTAGAG GTTGGCCAT TTTGTTCTT GATTTTACT GGTACTTATT TTGTACAAAT 840  
CAATCTTTT CTCCTCTG CTCCTCTG CTCCTCAC CTTGCTGCTT TCGCTCTC CATCTCTC 900  
TCCGCGCTC CCGCTCTG TCTGCTGCT CCGTCTCAT CTGTCACTC CATCTCTC 960  
CGCTCTGCT TCGCTCTG TCGCTCTG TCGCTCAC CAGCGCACT TCCCGAGTT GTCTTGGG 1020  
CTGCTTACT GTTCTAGTC CGAGCAGTT TCACTGAGG TTGTGAGTC GTGTTGGG 1080  
CTTTCCCAT CTGCTCTG TTTCTGTAGA TTGACGGTT TCTTTGTAT TTGATGTTT 1140  
CTGACAGAT TTAAAAAAA AAAAAAGAAA AAAAAAAA AAAAAA 1186

## (2) INFORMATION FOR SEQ ID NO: 90:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1821 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

## (D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 90:

5 AAGACATGCT TTCAAGCCCT CCCATATGTA TTCCGGGGCC CAGAGACAC TCAGCTGCA 60  
KATCTGCTCT CACTGCGCTC AAGATGCTT CAGGAGACAC CATGTACCA CCTGAAATTC 120  
ACATTCCTTC CCAGAGCTCT CCCAACATA GTCTGACCA GAGTGGGGC TCTCAGCTA 180  
GAGTTTTCAC ACACATGAC AGGCTGCTGG GGCACATTC AGAAGCCCTT TTCTTTCTCT 240  
CTCATGCTTA GAAGCAGCC CTAGAGCT GAGTTACTC CTTGTACTC AGCAGAGCC 300  
TGATTTACCA CAGCTGCCA CACAAAGCA GTGTATATAC ATCTGTTTAC CTTTCCCTAT 360  
CAGCCAGACA CAGCCCTCT TCCAGGTCA AACGACAGC GATGCACT CTGTTTTCAC 420  
AGTCAAATCA CTACTTCAT TGTACTTTTA GATCAGCAA AGTGTGACT CTCTCATGCT 480  
GTGCTATCC CTACAGGCC CAGCAGGG ATGCCAAG AGCCAGCTTC TCCAGGCTG 540  
CAGCAGAGC AACCCACCA GCTTAAATTC AGCAGAGAC CTCACACCA ATGTCTTCTT 600  
CTAATTAGAA GGGAGATT GGCACAGAA ATCACTTAT CTATATATAC AATATCTCT 660  
TACTCACTT TAAATCTCT ATGCACTCT ACTGCTTTTA AAGCTTATTT AAAAACTCA 720  
TACTAATACA ATGCACTCT TAAAGATT TGGAGGCC ATGCTGAGAA CAGTTCTCT 780  
GAGTTGAGA TTTTTCAGA AGCTACTTAC CTACATGTA ATGTGCTAT TTCTTTCTCT 840  
TTGTAGAGA ATCCCTCTT ACTTTTGA ACAGTAATGG CAGCTTCTAG TACAGCCCT 900  
ACAGTTTCT ATGAGAAAA TTAAAGATA CTATTAATTT GTTAAATAT CAAATATGG 960  
ATAATGATGG CCAGAGATT TACATACA AGTAATCTC AATGTAAAG TATTCAGCTC 1020  
TTCAAGTTG ATGCTCTGT AACGACCT GAGCTTCCAC ATCATCTTCA AAAAACTT 1080  
TCTCTGTTCC CCATGATCT CTATAGCT AACTCTTTAG TCCCTCATTT AGCAGCTTT 1140  
AATCTCTCA AAGATATCT ATCATGCTAT TATTTTACT TTACAAAA AAGTTGAT 1200  
GCGTTTAT TTTCATGCC TATAGCAGG TACCTTAGTA GGCAGATAT AGGAAAAA 1260  
AATGAGCA AACAAATCC TCTACAAATC CAGGCGAGA AAGTGTGCG CAGATGACT 1320  
CATCTCTCG TCCCTCTCAT CAGTCAAT CAGAGGCTG CAGTCAATC CTGTTCTTTG 1380  
AATGTGAGC AGTGTCTCT GTACTCTTT AAACTTGGC TATAGCTGT TTAGCACT 1440  
ACAGATTAAA GATACAGTTA CTTAAACAGC AAGTAATTT TATAGCTCT CATCTTTA 1500  
TCACTCTTG GTTGTCTAT TTTTCACT ACCTTTTCT ATCAGCTCT GTTCTTTTG 1560  
TACATTTTC TCATATTTGG GTTGCACAG TAAACACAA CTCTATTTTC AATGAAAAA 1620  
GTTATCTTGA TGGAAATTTA AACCAATTA ATTGTATTA GGTAAAAA AAAAAAAA 1680

AAAAAAAAA AAAAAAAAAA AAAAAAAAAA CTTGGGCGCG CAGCTTTT CTTTGTGGT 1740  
GAGGGTTTAT TTTGGCTTGG GGGAGTGGG CTTTGTCTTTT TACAGCTTG TGAGGGGGGG 1800  
AAGCCGGGGG GGGTTTCCCC C 1821

5

10 (2) INFORMATION FOR SEQ ID NO: 91:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 862 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15 (1) SEQUENCE DESCRIPTION: SEQ ID NO: 91:  
TGGCCCTTTT CCGACCGATT CCGGGCTTGG TGAAGTGGG AGATGTGAG TCCAAATTAG 60  
GAGCTGGAGA GAGGTGAGA ATTGTGAGG TGGCGATTG GATTGTGAT CTGACTTGT 120  
AATGAGTGG ACCTTGGCAT CTGTGTTCAA GTTCAGCGTT TGTGTGTGGG TTCTGTGGAG 180  
AGCTTACTCA CCGCGAGTC TTTTCTTCT CTCTCTCAA GAGAGCCCT GTTGTGCTTT 240  
TACAGCGCT TGAAGTCTCC CAGAGACA AAGAGCAGA CAGGAGCTG TAGGGAGAT 300  
TCTTTCTGT TTTCTGTGCT TTCTTTTGA CAGGCTGCC GAGAGCCAC TCATGGCAT 360  
GCCAGAGCT TTCTCAGAA CAGTCAATAA CAGTCTCTTG AGTCTCTTTC TTGTCTCCC 420  
AGCTGAGCTT TCTTATTTCA CTTTCTTGG TGTCTATAGG AATGCAATAG AAGACCTTGG 480  
GAGCTTTTTC TGTCTCTTTC TGGCCCTTCA TGAAGCCATG GGGCTGGGCC TTGGGGGCTC 540  
CTCAACCTCA CAAATTATTT CTCTCTCCC TCCAGCCCT TTTTGTGTGT CTGAAACCGG 600  
TTTAAATATG TGACTCTCCC AAGAGAGAG CCGCTGGCTG TATGAAACTT GAGCGGCTT 660  
TTGTAAAGTG CCACTCCCAA ACTTAAAGT AGCTAAACCA ATTTTAAAA GATTCATAG 720  
CTTGTCTATC CTCAGATGT AGCTATTGAT GTACACTTGG CAGCGAGTG TCTGAATATG 780  
TGTGTGCTCT GATTATAGG ATTCAATAT TAAATGTCT GCTGAATATA AAAAAAAAA 840  
AAMACTCA GGGGGGGGGG GT 862

20

25

30

35

40

45

50

(2) INFORMATION FOR SEQ ID NO: 92:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 696 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 92:

CTGAGCGAG TGAATGGAC TGTAGGGCT ACCGTACCG CCACTGCTGC GCGAGGGGG 60  
TGAAGGCGAG AGGGGGGGGG AGGGGGGGGT TGCMAAGATG GCTCAGAGCA GAGAGGGGG 120  
AAGCCCTTTC GCGAGGCCA GCGAGCTTGA CAACCCCTTT CAGGAGCCAG CTGTGATCCA 180  
GCGAGGAGCC AGCGGGGAT ATGCGAGCT TGAAGTCTAC AAGCTTTTTC AGAGCCGGGA 240  
GCCAGCACA GCTATGAGC CTGAGGCC TCGCCATATG CTTGAGCCCT CAGCTCTCTC 300  
CTTGAAGCC TCGAGAAAGC TGAAGCCCAAG AGAGCTTAAG AACTATGCT CATACAGAC 360  
TGAAGGCTCA GCTCAGCAG CCGAGCTGA GCTGCTGAG AAGCAGAGG AGCTCAACCG 420  
GAGGCGAGG GAGTTGAGC GAGGAGCCA GAGCTGAGC ATGCTGCCCT GGGGGGZACA 480  
GCTACTGAC AGAGCATATG GCGCCCTCTA CTTCTTTTTC GTCCAGTTCA GCGCTGCTTT 540  
TTCCAGACA TCTCATAGA GATGCCCAA GAATTTGAGA AGACTGTATC CACCATATAC 600  
TACCTCTGGA TGTGAGCAG GGTGATCTTT CTCTGATAT TGTGCGCTG CTTGCCAGAT 660  
TCTGTGTGGA AACACAAAT GCGAGGCTT TGGTT 696

5

10

15

20

25

(2) INFORMATION FOR SEQ ID NO: 93:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1886 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 93:

CAGGCGACTG AGCTTCTTT CCGAGGATG CAGAGGTCC TACAGAGAA GCGCTTCTT 60  
GCAATTCAGA GGGGCGACAG CTTGTACCC AGAGTCACC AAGCAGCTTT CTACTGGCT 120  
CTGAGCTTGG CCACTCCAG AGAGATCCA GAGGCTCTGG GTATATTCG CCAAGCTTT 180  
CAGCTTCAG GTAGCATGC CAGCTGCCCT CAGCTCTTTC CCGCTGCTCT GTACGACAG 240  
AAGCTTACC ATGAGCTCT GAGCATATC GAGTGGGCC TATGTGATA CCGAGAAAT 300  
TTCACTACTAC TTTTTCAAA AGTGAAGTG CAGTCACTCT GCGAGGCC GAGAGAGCA 360  
CTGCTGACTT GTAGGACAT CTTGCAATTA TGGAAATCT GTTACAACT CAGCAAGCCC 420  
AGTATTTCTG GAGTTGGAG CAGCTCTTA GATGAACCA TTGCTGACAG AGCAGACTT 480  
AATACATTA CTTTCCAGA CTTGAGGAT CCGAGAGAG GCTCTCTCA TCCCATATG 540  
GTACAGCTT CAGAGTGA GAGGZACTG TCGAAGTGG CTTCTCTCT CAGAGCATG 600  
CGCTTACCA GGGGCGGCTG CAGCCCTGGA TGAAGCTGG ACAGATCTG CTGATGACG 660

40

45

50

55

60

5 CTGAGCTCTA TATCGGCTAT GGGAGCGCTG CAGAGCGCAC AGCCTGTACC CAGAGAGCTG 720  
CCAACTCTTT CCCATCTCC CACATCTCC TCTACATCCG CGGCCAGATT GCTGAGCTCC 780  
GGGAGAGCAT GAGAGAGCG CGCGGCTGGT ATGAGAGGCG CTTAGCCATT CAGCCGCCAC 840  
CAGCTGAGGA GCATGAGCG ACTTGCGCTT GATCTCTCAC CAGTAGGCGC GTTACAGTTT 900  
GGCGGAGAG AGCTCTCGGG AGCGGTGCGA GGTGAGCTCG AGAGGCCAG AGGTCTGAAA 960  
CGCGCTGGGC GAGCTCTCC AGCTCAGCG CAACGATCCG CGCGCTACCG AGTCTCTCTT 1020  
GACAGCGCTG GAGCTGAGCG CAGAGCGCC CGCGGTGGCC TTACACATCA TCCCGCCGCT 1080  
GCTCTGAGCA GCGCGCTGCC AGCTCTACT GCGCTCAGC CTNCGAGGC CTGCGCGGCG 1140  
ACGAGGCTTT GTCCATCCG CCCAGGGGA TGAATCTGCC GCACTGAGGC CAGGGAGCAG 1200  
TGTTCAGTGG GGCACATCA ACCAGCGAA CGAACCCGGA ATCATGCTC TCGCCATGTC 1260  
GTTTCTCTTT GTTTTCTTTC CCAGCCCAT GTTAGTTTCT GAGCTATTC ACATCTCTCA 1320  
AATCGATCA TGTGCTATTT TTTGTTAGTT GACATCTGAG TTTTCAGTAA ATGATTTATG 1380  
GAATTAATCA GCATATGAG AGAATATAT TCAAGTTTAA AATTCAGTGG CAGCAGAGAT 1440  
TATTTTATTC AGAGCTGTA AGAATACAC TGTCTTTTTC TCCCGACAC CCGTCTGCC 1500  
CGACTTTTGG CCGAAGACA AATGTAGCT TCTGTCTCC CAGCTCAGCA CTAGTCCATG 1560  
CGAGGACAC AGCTGACAT TTTCTGTTT TACTGTGANT AATGTACCA TGTGATCAT 1620  
TACTGTCTTC ACTTAGACA AGCGTGAAT CCGAGATAT TTAATTTTAA CCAATATATG 1680  
CGCTGTACAA GAGAGGAAA TATGAGTTAT TTAGTTTAA CTTTCTTATG TGAATTCAGA 1740  
GTTTATTTAT CAGGGGAAT ATGTACAAAG AGCTTCAAA TGGATATTT ACCGACATTC 1800  
CTTATACATG ACAGACACTT GCTACATGG GAGATGATG TTAATATTA ATGATTTT 1860  
AATGGAAAA AAAAAAAAAA AAAAA 1886

(2) INFORMATION FOR SEQ ID NO: 94:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1774 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 94:

55 CTGAGCTACC GTTACAGTA GGAATACCC CAAATTCACA TGCATCTAC TGGAGCTTCC 60  
CTCTCTCCC CCCACATCA AGATGTTCTT TTTTCAATAC TATATACTAT TGCATTCAT 120

5 GATATCTCTG TAAATTAATT AATCCCTCAT GGAATGATA TTAGCTTCAT TATAGATAGA 180  
AGCTATATTA ACATTCCTGT ACATGTATTT TGTACTATGT GTGGTAATTT CTGTAGATG 240  
AATAACTAGA AATTAATGG ATCAGGTTTC ACATTTGAG TTTTGAANAC TACTAGCGAA 300  
AAGATTTCAC CAATTTAGAA CTCATCAT ATGAGAAATG CTTGTCTGCC TATAGTCTCC 360  
CAACCTCGAA TCCTTAANAA TTTTTCGCA TCTGGTAGGC AAATTTCTTT TCTTTCTTTT 420  
GAATATTAAT GAGGAGGAC ATCTTTTCT GTTCTCTGCC CATTTGCATT TCCATTAATG 480  
AATTCCTTTT GCGATTTTC CTTTCTTAA TATGAAAT CTATGACTA CTTCTCAT 540  
GTATTAANAA CACATCTCT TGAATAGAA GACCTTTTTC TCCATCTCA CCAATCAT 600  
TCCACTTACC ACAGTTTAC ATACATCTC TAGTCACCTT TCCGTAGCAA TATACATACA 660  
CATTAANACA CTTTCTTCAAT AATAGGATC TCAATTTCTG TAGCTTTTAA AAATTTTGGT 720  
CTCAANAAA CATTACAGCT CTTTAATTTT CTTTAATGCT TGAATATAT TAAATTAATAT 780  
GAATATGCA TTAATTAATC CTTTAATTTT TTTCTCTCC CATTACATTT CCAATAGTAA 840  
ACATCTTAT CAGATCTCTT TGTGATGTC TGTGAATTT TCTTTAGCT GAGTCTCAT 900  
AAGGTGAGAT TTTGATCAA AGGTTTGT TCTGTCCAC CTTCACTCT CCAAAAGGCC 960  
TTCAATAGTC TATTTTCAC AGTGTATGG AGAATGTCA TTTCCCATTA TACAGTACC 1020  
TACACTGTAT AGTTTATTC TGTGCGGCA AAAGAACT TTTCTTAAT TGCATTTGCC 1080  
TGAATTAATA AAAAAATGCT GAGATGGGG TTAATTTTCA GTTATTTGCC CATTATATAT 1140  
TTACTGTGGA TGTTTGTAT CCGTTACCTG CTTTCTATG GATTATATGT GGAATATATG 1200  
TTTTTAATG TCCACATCT CCGTCCCAT CTTCTGCTAA CACAACTTT ATTATTTTGT 1260  
GGGAACTTA TCCCTGTGG CTTAGGTGAG CATGTGACA GCGCTGGCT CCGAGTCCC 1320  
ACAGCTCTCT ACCACAGTG ATAAAGAT GGTATATTA CTTAGCCAG CTTAGCGAAA 1380  
GCGCTTACA GACTCTCTG TCGAACTACT GGAAGAGG CTTTATGAG ATCCGAGAA 1440  
CGAGGACA TGTAGCCCTG AATTTGTCC ATGTGGAG AGTCTGTCTG AGGAGAACT 1500  
CGATGCTGAG CAGAAATGGA AAGAGACTA AGTTCTGATG TCAATTTTCT CGAGGCCCTA 1560  
GATCCAGCTG TCCCTAAGC CTGCGCTACT CCGGACTTAA AATTTTGTG AGCANTAAA 1620  
GTCCCTTTCT TGTTTAGAT AATGAAATG AGTTCTGTAT CTGATTAATA TAGGTTATTT 1680  
GTATTTCTT ATTGATTTCT AGAAACTT TGTATTTTAA AATCTGAGC TTTATGACT 1740  
AATTAAGTAA ATAAATTAG CATGCGCTTC CATG 1774

(2) INFORMATION FOR SEQ ID NO: 95:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2503 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 95:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

GGGACGAGCG AAGGCGAGCG GGCACAGCTC CAGGACTGCA TCTGCTGCG ATTTCCTTTC  
CACTGCTGCT TTCTGCGATC TGACATGAGA AGCCACGCA GAGGAGAT TAAACACACC  
AACCTGANT TCTGCTTCTT CTTTTCATG AATTGTTCTG TGATCTCTGC ACTGCTTTTC  
TGTCTCCCGG CAGAGCGGAG TAGAGATGCG GGGCCGAGG CTTCTGTGCG GCGACGAGCT  
GCTTTTCATG AGCATATGAG TCTCTGTGAT TGTGTTGATC TGGCTGATGT TATAGAGCTT  
TCTCTGCGAG GCTGCGACCC TCACTGAGCT GCGCAAGCTG AGATCTGCT TCTATAGCTT  
CTGCTGTGCG AATGAGAGCA CAGACAGCTT ACATGTTGAC CAGTTCCTCTG AGCTGGAGGC  
CTTGGGGGCG CTTGCGGTTG GCTTGCGGCT GGCACAGCTT GCGTGTAGG GGTGCTGTGT  
GCTCAGCTTC TTGTCGCGCC AGCTCTGCTT CTTAGCGGAG TGAACAGATG ATGAGAGAGC  
GTGCGGAGTG GCAATGAGCT TCGTGGCTGT KTCCTCTGTC TTGAGCATGCG TGTGTTGCTT  
GCTCTGCTGC TGTATGTTGT GGAATGCTGTC AGGCTCTGCG TCCCGGGGCG TGCGTTTCTTA  
GCTCTGCGCA GCGCCGAGGC CTTACTGATC CTTCTGCTTA TAGCATGCGC TGTGTTGCTT  
CTGAGGCGTG AAGGCGCTGA GAGCAGCTT GAGAGCTGCT AAGGCTTAC GTGATTCGAA  
GGTTTCAGTT CCAACATGAG TCAAGAGTGG CACATCTGCT CAGGCACTTC ATTATTAGAG  
TAAAGCTGAT CTCAGCTGCC AGCATGCGAA CCACTAGAG AGGAGGTGCG GCGCTGTGT  
CAAGAGAGGC GAGGGGAGGC AAGGGCAGAC AGGGCAGCTG TGACTTCTTA GTACAGATTT  
GTCTGTGCTT CAGGAGTTC AAGGCTGCGA AAGAGTCTCT AAGCATGCGA GCTCATGCTC  
ACACCAATTC CTCCTTTTAT TATGTGATCT GAGCAATCTT TCTCTAGCT TCAAGAGGCT  
GGGAGGCGAG TGATGCTGT CHTGCGGCGA GACTTTCGCG CTGATTTGCC AATATCGAAA  
ATGAAGACTA GGAAGACT TACGCGAACA AACGAGGACA TTAAAGAGC GAGCAGCTCA  
GTGTCTCTCG GAGATGCTT AAGGAGCTTC CACTCAGCCC ACATATGTA GTGGGCGGCG  
ATAGGCGATC ACTGAGACT CAACTCGAGA GGTTCAGGAG TGATCTCTGA GTGACTCAAC  
AAGAGCGGA CAGTGGGCTT AAGAGACA GGTTTGACTG CTTGAGGCT TCGCTGGAGC  
TGAAGCGAGA CAGGCGAGAG GGTTCGCGTG ACAATTAAT CCAATGATCA GACCTCGAG  
GACTCGAAT CTTGCTGTG AAGGAGCTG GAGGTTTGG CAGAACAAA CCGTGGCAGC  
CTCCAGTCC CAACTCGAGA CTTGAGAGA CATTAGACA ACTTAGGAT TGGGATTTG

1620  
1680  
1740  
1800  
1860  
1920  
1980  
2040  
2100  
2160  
2220  
2280  
2340  
2400  
2460  
2503

TGATGATTTT CTAGCACTTG TGTATTCGAA AACCTGATG GCAATGATTT ATTCAATAT  
TCTGTTCGAA AGGCACTG AMAGAGAG CAGAGAGATG TACTCTGCTG TCAATCTCTG  
TCTCAAGTGT CTTCTCTGCG CTTCTGTGCG TCTGTCTTGA TAGCTAGCTG CCGCGGAGC  
AAGGTACAGG TGAAGCGAG GTAGAGCTTT GCGGAGAGAG GCTCTCTCTG CTTACAGATC  
TATACACTGT GGCCTCAACC TCCAGAGAG GCGAGAGAC TGTGGGAGC TCGTTTGTCT  
TCTAGGCTCG CTGAGAGGT GGGAGCTGAT TGAATGACTC ATGATGAAA CTAATTTTGA  
AACAGGCTTC CTTCTTCAGG AGATGATGAG GCGACTTAAC TGTAGCAATT CCAATCGACC  
TGGAGTGT CTTTCTTCTT GCAATGACTT GTCTCTTTG TTTCAATGAG TTGAGAGCAGC  
ACATGACATY ATTTTGGCTG GAACTGCTG ACTGACTAAC ACAGACATTT GCGAGCTCGAG  
AGCTTCAGGA CCGAGAGAG GCGAGATGAC ATAGAGAGG AATCAATGAG GATCTCATTT  
TGATCTCTG CTTGCGAACA CTCAGAACCG CAACTCGAG CCGCTGAGTG TCCAGAGGCTT  
CTGGCTGAG GTGATCTG CAGGCGCGAG AAGCGAACA GTTAGAGAGC CAGAGAGGCG  
CAATTCGACA GCGGCGGCG CAGAGAGAC CAGTGGAGC TCACTCTGCG TCCAGCTGCT  
CTGCGAGAC AATTCTATG ACATGTTATG CCACTAGCTGA AAGAGATCTC CCGAGATGAT  
TTACAGCAT CTTCTCTGCT CTTCTCTGCA GCGCTCTGCT TACAGCGAGA GCTGAGACCC  
CGCGGCTCT GGGAGGCTT GAGGCTGGA GTGATGCTT GTC

(2) INFORMATION FOR SEQ ID NO: 96:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2801 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 96:

CTGGAAGCC GAGGTTAGCG GAGCGGGCG GCGCTCTGAG AGCGCGGCT GCTGCGGCTG  
CGTTCGCTC CCGAGAGAG ACCGAGGAG CAGAGCGGCG CCGCTCGGCG TCTCTCTGCA  
TGAGCGCGA GTGAGCGCG GCGCTATG CCGAGCGCG GCGCTTCC CCGCGTCTT  
ATCGAGAGC CAGAGAGAG GCGCTCTG CAGAGAGCG GAGAGAGAG AGCATGTTG  
ACGTTTGA TCGCGCGCA GGTGCTGCT GCGGCGGAG CCGCGGCTG GCGCGGAG  
GGGCGGCGC TTAGGCGGCG GTTTTTCGA AAGGAGCGR GCTCTCTGAG CCGCGCGCG  
ACGAGCGCC GCACTCGAA GCGCGGCTT TCTGCAAGC AAGCGCGCTG CCGCAAGCCA

60  
120  
180  
240  
300  
360  
420  
480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560

5 GCGAGCCCC GCGCGCGGG GCGCGCTGG AGGTCCCGCG CAGCGCCGAG CCGCGTCCG 480  
GCGCGCGGGC GCTCCGAGAG CAAACGAGGC CCGTGAGAGC TCCACCTAGT TCCACGATA 540  
AAATCCGACA GCGAGCTGG GAGTCAGAA TGGCTTAGGC CAGGTGGTGT GTAGCTCTGG 600  
TATTATGCTT TAACTGTCT GTGATGGCC CTGAAATTTA CCGTTCAAGT TATCTTCCA 660  
GTTACAGAG ATCCATAGAG GATGGTGTG AGGATATACC TACTCTATCA GAATATGTTT 720  
AGGATTTTTT GAATCATCTT ACAGACAGC CTGGCACTTT TGAAGCTGAA ATTGACAGT 780  
TTGCGAGAG CCGTGAATGT TGTGTACAA CAGTGAATGC TTTCAGAA TTCTGGAAAC 840  
TCATCATCA ACAGCGACA TCTATCCCA ATTCTCTCTA TATGGAGCT CCGCTGTGTA 900  
ATTACTGTC CATTATCTG ACATTTAGC CAGCAATGG CAACTTCCG CAAATGCTAC 960  
TTCAAGAGT TCGGACTGAA TATGAGTTA AGNTCAAGC TCGAAGGG GATGAGTTA 1020  
CTCGAAGAG ATTTCATCA TTGTACTCT TTCTGGAGA ACTTTATCTT AACCTGAGA 1080  
TCAGGAGAC AATGGAGAG GTTACAGAG CAGATATCT TCGGTTGGT CTTGCGAAT 1140  
TCTGAAATC CCGTTTCTT AATCTATGG ATGACATTT AATTGTGCA GTAAATTTGT 1200  
TAAAGTTGAC AGATCACTT TTGGAGTGT CTTGGAGA AAGAGNAG ATGATATGG 1260  
AAGAAATAT TCGAGAAAT GAAGAGTTT TCTTAGATC AACTGCAAT AGAGATGTA 1320  
AACAGATCT CTTGAGACT GTAGACTCC GGTCAAGTAA CTGGCGACA GTCCATGCA 1380  
CTTCACATA TAGAGAGCA ACACAGAA ATGATCTTAA CTACTTTATG ATGACACAA 1440  
CAATTTATAG ATCTGATGT GTTCTTTCA CTGAGCTGA TCCAGATTAC CAGAGAAAT 1500  
ACCAGAAAT ACTTGAAAG GAGGACTTTT TTCCAGATTA TGAGAAAT GGAACAGAT 1560  
TATCCGCGGC TGGTGATCCA TACTTGATG ATATGATGA TGAGATGAC CCAAGATAG 1620  
AAGAGCTTA TGAAAGTTT TGTGTGAT CAGAGCTTAA GCGAAGAG TGAAATGAA 1680  
TTTCACATA TCAATTTTAT AAGCAGTTT AGTATGCTG ATTGAGCA ACACAGAGA 1740  
CGAGAAAT GTGTCAATC TATACCAAT TTAGATGTT GAGTTATGTT ACTATGTTAT 1800  
GCACTTTAA TTTTGTTAA CACTATCTG CAAATTAAC TTTATTCCT ATACTTAA 1860  
ATGTATAT ATATATATA GTTATATG TACAGTTAT TCTACTGTTT TGGCTGAA 1920  
AAATCGAT TTGAATGAA TGAAATGTTG AAAATTTGCG TACTTGTTTA GATGCTATC 1980  
CTTAAATTC TACTTTCTT GAGGGGAAA AGTCTGTGTC TGGAAATACA TATPACTGA 2040  
AAATGTAGC ATCCTTTTT AGTATGAGT ATTATAGCTT YCATTTTAT TKGACATTTA 2100  
GTGTCCCAT GAATGAAAT TCAATATGA ATCATATCT TGAATATCTT TAGCACTTAA 2160  
GTCTTGGAA TATATCAAC ACTGATTTAC ATATGAGAT CTAATTTTGA TACCAAGGC 2220

TTTTTAATG TCAATGGGG GAAAAACCA ACTTGTGGA ACTCCAGCT AAACAACCA 2280  
GACTCAGCT GAGATTTAT TCAATCTTA GGAATGTC TTTTATAT TATTTTTC 2340  
AACTGCTTA CTTCATACC TTAAGGCTA GAAATATAT CTGCTTAT TATATGCTT 2400  
TCTCACTTT ATTTGTAGC AAGGTGCA TCGACTTTT TACTAGAA TTTTACTGA 2460  
TATTTGAT TCAATTTTC ATCTCTTAA TAAATGATC ACCTTAGGG TCACTTTCT 2520  
AATACTTTA CTATATGTC GTACAGCTC AGCCTAATA AATATATTT TTAACATAG 2580  
TCAATCTTT TTCAAGTAA CTAAAACTG TATACAAAG GATTCCTCT AATATGCTAT 2640  
GTAAATGTT CTATATTA CCACTGTTT TCAATTTGT ACATCTGCT CTGCTATATC 2700  
AGTATGTT CTCACTTTC TCGTTGTTG TTCAGTCTGA ATTAAATTA GACTTTGAA 2760  
ATTAAGCTTA AAAAAAATA AAAAAAATA AAAAACTGA G 2801

## (2) INFORMATION FOR SEQ ID NO: 97:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1631 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 97:

ATGAGCGAA AGCAATCAC TGAATCTTGG GTTCTTAAATA TAAATTAAGG TGTGCTGCT 60  
AATTTTCTC CATACATAT CATGCTATC AGTGAATGCTC CAGTGAATGA ACTGTCCCTC 120  
GAGCTGCTTC TCGTTCAAGT TGTCTTCCA GCAATTAAGT ACAGAGGACA CAGAGGCGAG 180  
TGGCTGAAGG GCGTGTGCG AGCCTGACT GTGACCGCGG GATTAATGCT GCACTTCTCAT 240  
TCTTATTTAT TCGAGACCA GGAAGAAAT GAAACATG GAAATCTACA AATTATGAT 300  
ATTCAGCATG CTGAAATTA CAGGCTATT CCGTGTGTTG GAGAGGCGCT TCAATGAGCC 360  
CAGCAAGCCA TACTTCAGCA GGAAGGCGCT GTTGGTTTC AGCTTTACCG CCGAGCTTTA 420  
AATTTTCAG TGAAGATAT TCTGTGAAT GTCTTCAAT GTATTAAGAT ACTGAATGCC 480  
AGCCTCATCT CCGTACTTT ACCGATATT CCGTGGCGCT GTTAAATGTC GTTTTGAGCG 540  
GGAGCTGCCA AAATCCATA GCTCTACCA CCGTGTGTTG GTCTCTATGT TTGCTGCTTA 600  
ACCAATGAG CTGTGAGGCT GATGTGACA TGAATGCTC AGGAGCGAG AGTGAATCTC 660  
CAGAGGTTA AAGATGCTC TCTATGATC ATGAAGACTT TGAATATGTC GGTGCTGTGG 720  
GCTGAGTTG TCCCTCTGCT TCTGGGCTC CTGTTGAGC TGTGATTTGT GGTCTCCCTG 780  
AGGTTGCTT TGAATCAGAC TCTCTTTTT TATCATGAC AGGACTGGCC ACTTGAGTC 840



320

321

5 CTGATGCCA AATCATCTGC AGCTATACCA TTGATGGGTC CTCAGTGGTG GTTGAAAACT 900  
GTATTTGAC AGTTTACGC AATGGGATC GGGAACTTG ACCTTCACTA TATTTCTGT 960  
AAACTGGCAG CTCGCGTGT CTATGCTGTC TTGCTTTGCC TGTGTGTAC TTATGTCTA 1020  
GGTTCTGGTG TTGTCTCTTT ACTAGTGTCT ACTGGGAAA TGCAAAACCT AGTCCTGCG 1080  
CGAATTATC CAATTATCT CATGCTGGTG GTATTGATGG CAATTTTGTC CTTCGATGC 1140  
GGCAGTTTA AGCGCTTTA TGAACATAT AAAAATGCA AGTACTCTGT GGGTCAAGCA 1200  
CTGTGAACT AGAAGGAAA ATCTGGAAA CAAAGGCTAT CTCACACAC TCCACATCTA 1260  
TCCAGAAAT AAGATGTTG TCTCAACAC TTGACTTCC CCTTTACATG TCTTTTGTG 1320  
TGGACTCTC TCTTTGAGCA TTATGCTCAG TGAATCTCA GCGTTGTATTT TAAATTTAAT 1380  
GTATTTGACT TGTGTCTCA GATTCAGAG AGCAGCGGTG TAAGATTTCTG CTGTCTGCC 1440  
TGGATCTCTT GACATTACTG CTGTCTGAGA TTGTGTATATG TGTAAATACA AGTTCTCTGA 1500  
TAACCTAATA CTTTGATTA AACAGATGT GCATTTGACA TCTTTAACA AATGTATAT 1560  
TATTTATTA ATCTATGTTG TCACTTTTAA AAAAAAAAAA AAGGGGGGCC 1620  
GTTACCAAA T 1631

(2) INFORMATION FOR SEQ ID NO: 98:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 504 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 98:

40 CCGAGCTGG CAGAAATGAG GGGAGGAC GAGCGGCC GGTGGCGTT GCTATGCTT 60  
CCAGAACTT ACTGAGCAG CCAGCTGAGA AGATTTGAG GAAATGCTG CTCCTGGTC 120  
TCCAGACCG ATGATATAG TCCAGCCAAA AATAAACTT CCGCCCTTCT GCTTCAATG 180  
GAAAGCCAC GTGAGATGC TGGCTGTA TATATACAC TCACTGTATA CACAGATAT 240  
CATCTCATC GTATCTGTGT TGGCACTAT ACCAGAAACC ACAATATCA CAGTTGGTG 300  
AGCGGTGTTT GCACTGTGA CAGCATATG CTGTCTTGG CAGCGGCC TTATTTACG 360  
GAAGCTCTG TTCAATCCA GCGCTCTTA CAGAAATAG CTGTGATG AAAAAAAAAA 420  
AGTTTGTAA TTTTATATA CTTTATATG TGAATATAG TATTAACAT ATTTCTGTAT 480  
TCTTCCAAA AAAAAAAAAA AAA 504

(2) INFORMATION FOR SEQ ID NO: 99:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1416 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 99:

5 GGCAGAGG AGGAGGCTT CTCGTTGGG TCACTCTTGT GTGCCCTTTA GACAGGCTGG 60  
CTGCGCGGTT CCACAGGATA CAATTAGGAC TTGAATCTTT CTTTTCCTGT TTGAGTTTGG 120  
TGAATGATG ATAGGATTAAC ATGGCGCTTC AGGATGACCC CTTGGAACTG TGGCGAGTTC 180  
CTTAATATCT AGCTGGGATC CTGACCTTGG GAGGCGCCCTG TAGGGCCGAG CTCTGGAAAA 240  
ACCTGGGAT TGAATGCCGA GCTGTGGAG AACTGTGCTC GAGGCGAGGG TGGCTCTGAA 300  
CACTGATGT TCTGGGGCTG GAGGAGGAG GGGCTCCGCG TTTCCTCTGA ATGAACACTG 360  
CTCTTCAACA GTTCAAGTAC TTGTCTCAA AACTTTTCT AATTGATTCG TAGGTTTTCA 420  
TAAGCATTT TCTTTAAGG CATGGAAAG GAAATATCT CAGCAAGTC ATGTTTTGTT 480  
TCAATGGGAT GGGCGCGGCT TCTACTGCT GGGCGCTTCC CTTCAATGTC GCACCTTTGT 540  
CGAGGGCCA CAGGCGAAC TCTTCCCACT TTCTCCCACT GAGGACCAA GGGGCTTGG 600  
ACCTAATTT GCTAATACAG AGGCAATTTT TTGTGCTGAG TATCTTTTCA ACTGTCTCAA 660  
CGCTCTATT TTTTAAAG TTCTGTGCT TGTATTACA CAAACTAGA GAGAAATAGT 720  
TTCTGAAGCC AGTTTATTTT GAAATATCCC AAGGGAGGT TCGGTAGACA AATAATATTA 780  
GCTGCTTAG AACTGACA GGCAGACAG CAGAGACCA TTGGAGACA ATTTGCCAAA 840  
GATCTACCT GAGATACGC CTGTCAATG TCTTCAACAC GTGAATTAAC AGCGCTCAA 900  
AGTGTTTTTC TCTTTTAAA AAAAAATTC CACAGCTTTT TAAGGTGCA TTTAAGATTC 960  
CATGTGACTT TAGAATGAAA CTGCGCGGCC TGGCACTGT CAGGTGTCT AGAAGGTTCG 1020  
ATGCTCTGG AATGCATGTC ATACTCATCT CAATTGTGTT TCTTTAATG CAATTTTGT 1080  
CTTTTACAG ATCTGTGCT GTGGTGTGT TCTAAGAAAT CCGACACTT GGTTTTGTG 1140  
TTAATTTGAG CTGGCACTT GAAATGAGT TCTTTATATG CAATTTAGGC ACGACCATC 1200  
TGTGTGCTT GGTGTGCTG TATGAAATG AGGGAGGGA GGGATGTGAC CCAAAAGTAA 1260  
GGCGTCCCA TTGGCTTTGG CAGGCGGAGA CACTTCAAT GATTTACATG GTTCTGTGTA 1320  
AATTAAAGT TTAATGTAT AAGCGAGC TGTTCCTGTC AACTGTATA TTTTGTAAAT 1380  
AATATATG CTACTTGA AAAAAAAAAA AAAA 1416

## 5 (2) INFORMATION FOR SEQ ID NO: 100:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2847 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 100:

15 GGTATGAGCA ATTGTGTGC TTATCTATC TCTGCAGGA CTGGAGATT TGGCATCCA 60  
TTATTTACA CACCAATCA TATCAACA AAGTACCTA AAGAAGGAC CAGTGGCAC 120  
TCTCGAANA ATTATGAT CAGAGATTA AAGAATTTT AGATTTTGA AGCTGTATT 180  
GTCTTTCCC AATATCATT GTTTGATCT CAAATGATG CTTTATATA GCATATGACA 240  
GATCATGGT TCTCCATC TCAATCATG TTACTACTT GATATCAGTA TTTCGGCMA 300  
TTCAAGCATT TATCGATGG ATATTAATG AATATPAAA ATATTTCCA ACTGTCTCA 360  
GTACTTATC ATATCTGT GATCTCTCA GGAAGCAGT TTTCCTTTT CTTAGAGAC 420  
GTTTGAGATT TCGTTTATG ACTCTGCTG TCTTCAGTAC CTGATPAAAC TTTAACGAG 480  
GAGCATTAA ACACATGCA GCACCTTTG CCNAGCTTC TAAATCTCTG CCGGACGAT 540  
TTATCAATAT AAGAACTAG ATGCTTCTG CAGTGGCACT ACCTTCCCT AGAGCTGAG 600  
CATCTGCTT GCGCTTAGC CCGAGATGA TGAAGCTTCC CTCCTGGCAG GTCAATPAAA 660  
GTTAGAGAC TCAGATPAG GTCTTGGCTG GGTCAAGGTG GAGGGTTTG CTGAAAACCC 720  
TAAAGAGAG TCACCAAGG AGGCAAGTAA TGAATGTTT CAGATCTGT CTGATCTCA 780  
TAGCAATTC TGGATATCT TCAATGTTG AATTTCTGA TCGTGAAGG GACTTTGAT 840  
TGATATCAT AATTCAGGA CAGTCCGAG AAGTGGTTG AGTCTGGCT CTGACGCC 900  
AGAAGGGAA ATACTGTGA TTAAAGACA ACTATGACC AGGCCCTGAG CTGTCTCTTA 960  
GATATPAAA CAGATGGGA GTGGAGAGT CAATTGCTC AATTTATGA CTGAGGAAT 1020  
ACTAAGCCA AATCTGAAC GCACCTCCC CTAAATCTGT GACAGGCAC TTGTATCAC 1080  
ACACCATGT CCACTPAAA ACAGAGAT AAGAAGCTT CAGGTTTTCC CACTGTGTC 1140  
TGACATCCC AATTATGAA TCTTCTTCA AATGATTTT CACATTTATA GTTAGGGCTC 1200  
AGAATGCGA TTGATGAGC CTTATTTCT CCCTTAGCA GATGCTTTAA GTACATGTTG 1260  
CTGACTGAG CCAACCCCA GGAATGAGA GAACATTTCC TTTTCAATC CATCTTCAT 1320  
AATATGAGT TTTCTGGC TTCAGATA TAGACTTTG CACGATGAT AATATGAG 1380

5 GGTGTTCTG TCTACTCA ACTTATTTG AATATCTG CAGCTCACT CCGTAGAAA 1440  
AGAAATCTT CATATTTTAT TAACTTAGC CCGCAATGTA CTCTGTAGG ATGTGCAAT 1500  
TAAAGTCCA GTGAATCTG CTCTTACT GATTCCTGT TTTATGTGT GTTCTGGGG 1560  
AGTGTGACC TATATPAAA GGAAGGTGT CATATGTGT TATATGTATA TACATANTA 1620  
CAATGTCCA CACACACA CATATTTGA GAGTAGAGA AACTCMAAG CAGCCCTTC 1680  
ATTATCTGCT GTACTACTT AAGATTTCT GTGAGCCTA ATTACAGCT TCAACATATA 1740  
GTTGGGCTT AGTACTTGC TTACAGGAG AGCATTTCC TACCAAGGT CATTAAGTCC 1800  
TAAGGACTG AUTCANAATG ACGCCTTGA AGAATTTGC ACTCCAGCC TCTTCAGGA 1860  
TGTCTATTA GATGGAAAC TTGGATCCC AOCATTTTG GTGACTGTAG AATCTACTA 1920  
CTCAGTTAG ACTTAAGGC ACAATGCG AATTCATGAC CTGTAGTTG TGGAGGCTC 1980  
TAGAGTCC TCTCTGCCA AATAGAAAT ATTCCTTGC CATCTCTGA ATTCAATAT 2040  
CATATATG CTGTGCAATA CATCTCTC AATAGAAA TTAACTGAT GTTTACTGTG 2100  
TCTGTATCA ATCAGATTTT TATGTTAAA AATTTCTAT TATGATTTGA GTCCAGGCA 2160  
GCTTAGAG AAAAGAGG CCGATTTGG AGACTCATG CTCATTTATA TTGCTTTAT 2220  
CGAGCTGC TTATPAGCC CCGTACCTG TCCATTTCA GAAACATTA GACTCAGCA 2280  
GTTCTGATT CTGAGGCTT GCGTGTAG ATAGATGAT ATAAATTTGA ACTGAGACA 2340  
TACAGAAC ACAGAGGA GGGCAGGC AGAANAATG AATATATG AAGACTTAT 2400  
GATATATG GTTCANAAA AGCCAGGCG CCGTACTG GCTTGAATG ACTTTAGGG 2460  
GACAGGAT TAACTTTA TGAAGGTTA ACCTGACCG TCTTTCTCA GTGAGGCC 2520  
ACATGCTG AATGGGAGA ACGATGAT CAAATGCTT CAGCTTCTT TCTGTGAC 2580  
AGTCAATTC CCGCTTAGG AATCTTCCC TTCCAGCTT TACATTAAC AAGGAGAC 2640  
TGAATCTTC AAGGAATTA CAGTTTGG TTAAATTTT AGTATATCAT TTTCAATG 2700  
TAAATTTT TGTAGAGAG ATTACTGT ATCCAGAT GTTCCAGCTT GATCCCTGT 2760  
TCAATTTGA AATCATAAA CTATTAAGT AATTCGAA AAAAAAAAA AAAAAAAAA 2820  
NAAAAAAAA GAGGGGGCC CATTACC 2847

## 55 (2) INFORMATION FOR SEQ ID NO: 101:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1394 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 101:

5 GAGATTGAG GAGAGAGATA AATATCTAG AAGCAAGAT TCACTAGAG CAGAGAGCA 60  
CCCCAGAAA AAGATATGAA GCTATCTAT CTTCTTTTACA AAGAGTGGCC ATGACTTACT 120  
GTTCAGAACT ACTGAGATGTA TATTATATGT TCAATTTGCA ATTATATGTA CAGAGAGAAA 180  
10 GAAATATTTT ACTCTGTGCG TTATTCTACT TGTCTGAAAAG CTGTGAGACA AATATATAG 240  
AATAGACAA GCGCTTTTCT TTGTATATTC TCGTTTCAAT GCAATATATTT TTATTATACC 300  
ATTATTTTCA AAGAGTTTGG CATTGCGGCT TCGTACTTTT TTCTTTTATC GAGCGCTTTC 360  
CTTTTTTTGA AAGAGAGGTTTA TATATATGAG TTGATATGAG AAGTCAAGTG AAGCTTAAAT 420  
AAGAGAGGCA GATAGAGGCAAG TTACTTTTAA AAGACTTTAT TTCTTTTGAAG CATTCTTGA 480  
20 AAAAAAATAG GAGTGGCACT GCGTTTGAAT TCGCAATATG TTGAGAGGCA CTATATGAGG 540  
TTGAAATATG AATATATATTT TCGCAATGAG GAAAGAGAAA TTCTCTTTAG AAGAGTGGCA 600  
AATAGCTTTG CCGAGTGTCC CTATTTTATAG GATCTTTTCC TTCTTATATC CTTCAGATCA 660  
25 GAGTGTGTCC TATACAAAG TTCCCAATCA TTCTCTCTCA TATTCGCCAT TTATTAATGA 720  
TCACTTCTCT TTCTTTTACC CTTTTCCTGT TCAATATCAT AAGAGATTTG CAGAGTTAGG 780  
30 ATCTATCATG CAATATGCCC TTGTTCATCT GTGTCTCTTG CAATCTATGC TCAATGAGAA 840  
TTCTGAGATG CAGAGAGGTA GCTGAAGTTT GAGTGTGAGA CTGAGATATG GCGATTATGCC 900  
ATCTATGATA TTCTAGCTCC CTTCAGACAA GCGCGATCTT GCTAGTGTGC AAGAGAGCAAA 960  
35 GCGTACTGCC TTGTGAGGCTC AATTTTCCCTT CCGCTTATG AATATGAAAAG AATATCTATT 1020  
TTTCTGTGAG GTCTAGCAAT GCTGAGACAA AAGTATATGC ATTATGTATG TTATTTGAGTGA 1080  
40 TTGTGTGAAA ACTGAGAGAG CTCACTGCGCT ATTAAGAGAA ATTAAGAGAG AAGTGTAGAG 1140  
GAGAGACAAA AAGAGTATTT ATTGTGATTA GATTCAGACA TCCCAACTCT TCGTCTCTCA 1200  
GTCCCTGTCT CTGATTTTTC TTGATTTGATG AATCTTTTGT GCGACAGACC AATATCTTTT 1260  
GCAATCTGAG AATCTGTGAA GAGAGATATTT GATTCAGACA GTTGTGTCTCA TTGTTTTTTT 1320  
GCAATCTTCT TTAAATATTA AAAAAAATAA AATTTATAG TTATATCTTA AAAAAAATAA 1380  
50 AAAAAAATAA ACCC 1394

55 (2) INFORMATION FOR SEQ ID NO: 102:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 794 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 102:

5 GAGAGAGAG GAGATATAGAG GACTATAGAG AAGCAATGCG GCAATATATC TTATTTCCCT 60  
CGCTCTCTCC GCGCGCGTAT CTCTTTTTCAG CCGTTCTTCCA CCGTCTCTAG GATACATATG 120  
10 CCGAGAGTGA GCGGCACTAC AGTTCATATC GATCTCTGAG TCGTCTTTGAG GAGCTGAGCC 180  
GTCCAGAGCCC GAGGAGGAGAG GAGAGCCAGAG AAGCTTGGCC GCGCTGAGGGA GAAAGAGCTG 240  
CAGCTCTCTC TTGTGAGGTCAG AAGATCTGAT TTGTGTGAGA GATGTGAGAA AAGACTGAGTT 300  
15 TTCTTTTGA GAGCTGATAT ATCTGCTTTT CCGTGGCAAG TTTCATATGTC ATCAATATGAG 360  
GTTCCTTATC TCAATCTGAGC TCTTCTCTCT GTACAGATCA TCTTCAGAT CTATAGATGTC 420  
20 GTATTCAGAG CTGTACAGAA TCTAGATATA GCGCATATCA TCGAATGCTT AACTGTAGAT 480  
AAGATATCTCT CTGTCTCTAG AAGCTTTTCA TTATATATATG AATGTGTGCA TGTGTGCAAT 540  
CAGAGAGGAG CTGAAATCTCA TTATGTGTCT CTTTCTGTGTA GAGATATCTAG TTGACTCTCT 600  
25 GAAAGTGTCT GTTCTTATCT GCGTATATAC CTATATGTGAT GCGTATTTTA ACGAATATAC 660  
CGCTTATAT CTGTGTGAGC TGTCTATTTT CAGTGTGCGG AATGTATATG AAGATATCAA 720  
30 GAGCGATAT GATCTATATG TTGCAATGCG CCGAGATATG ACGAATATCA TTGTGAAA 780  
GATCCCAAGC AAAA 794

(2) INFORMATION FOR SEQ ID NO: 103:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1544 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 103:

45 TTGTCTTCT AGTGTATACC AAAAAATGT TTGAGCAATT GCTGTGTGAG CCAATTTTGA 60  
AAGAAAGAGG TTGTCTTCTT CATTGCCCCA GCGAGCGAGC TTGTCTGTGAG CAGAGCTTTC 120  
50 CTGAGTGAAG AAGAGATATC CTTTTCTTGG GCGCTTGAAG TTGCAATATG AAGCAATATG 180  
GCGAGAGCTG CCGTCTATGC TTGGGCTTCAG CCGAGCTGAG GATCTGCTGAG GATCACTTTC 240  
TTTGAGCTTT AAGCAATCTC TTGTGATATG GATATGATTT GCGAGCTTTC AAGTATATTC 300  
55 AAGTATATGAG AATGTGTGAG TTGAGGTATCA CTCACTCTTC CTTCGCAACT CTAGCAAGCT 360  
TTAAAGATATG TTCTTTTGGC GCGGCTTGA GCGAGAGAGC TGAAGTGAAG CTTGAGTATGA 420  
60 CTTCAGAGAG GAGAGAGAGG CTGAGCTTCA GATGTGACA CCGAATCTG GCGAGGCTTT 480

5 CAGAGGCGCT GTGTGGGGCC CCCAGGATC CTTAGCTGAA GCGGGGAGAC TCACTCTCCA 540  
TCTCAGGAAA TTCTAGCGCTT TCCCTCTAGG GAGCCACGGT TGAAGGTTGAG CCCCAACCC 600  
TCCCTTAGGG CCGCTGGGTTGG GCAAGTCTGG CCGCTGGGGT AGGAGAGGGAG ACTCAGGGCC 660  
ACACTGGGT ATTCTTCAAT TTGAGAGAAA CACAGACTCA CCGCGCACTC ACTGATTCCT 720  
ACACATGGCC AAGATTTCAC ACATGTGACC AGGGGCGACC AAGTGCCTTG TGAACCTTGT 780  
GACTAGATC CTAACTTCTC TATTCTCTCC TGGGTGCGTG GGTCTGTGTG ACCTGGGGCA 840  
GTGTGATAA TGTATTATTC TGTGACATG TTTTATTGGG GTGGCACTG GTTCTCCGAT 900  
CGCTGGCGTG GTGTGAGCC CAGGACTGTA GTGCTGGGAG CAGTAAGCT CAGCTCTGTG 960  
TAATGATGA TGTATGTGCT TGTCTGTGTC TTATGATCCA ATCTTTTCTT ACATGAGCCC 1020  
TGTGTTGTT TTATGGCTAG TCTTATCTGG CCGTGTATTT TCTTTCGGG GAGGAGGGG 1080  
TTTCTAATC TGTCTCCAGC CCACTTATT ACACACCCAC CTGCTGGGA CTTACTGCTC 1140  
GGGAGCGAC AGACAGGGAG CCGCAGAG TGCGTTCCTG CCGCTGTGCT GGGGGTGGG 1200  
GGAACTGGG GGCACATGTS GCGCTTGCTT TCTGAGAGC TCTCAGTGCC AGGCGTTTGA 1260  
GACTTTCCA CATGTATAAA GAAAGGGAG GTACAGAGT TCCATTTGCC TTTTATTTT 1320  
GCTGGTTGCT ATCTGTAAAT GTTTAATAA TATCTGACA TGTATCTATC AAGCCGACGA 1380  
ATTTCUAGT CTCCTTCAC ATATGAGCG TTTTAGGATG TTTATATTC TTCAATCTTC 1440  
TTGTTTCCA GGTTTTCCAG GCAAAAAG TCTGGAAITA TAGATACAGC TTATATTATA 1500  
ATTGTTCTT GCATATAAA AAAAAAAA AAGCAGGGG GGG 1544

## (2) INFORMATION FOR SEQ ID NO: 104:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 871 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 104:

50 ACCACGGCT CGAGCTGTG CACCCGCGG CTGCGAGTG AGGTACCAGA TTCAGCCAT 60  
TTGGCCCCGA CGCTCTGTGT CTGGAAATCC GGGTCTGGG GATTGAGTTC CCGGTTCTTA 120  
AGGTGGGTCG CTGTCCACCC GGGGGCTGG GAGTAGGTTA CCGATTTGAG CCGATTTGCG 180  
CCCGACGGCT CTGTCTCGG AATCCGGGTS CTGGGATTTG AGGTCCCGGT TCTTACGGGA 240  
CTGCAATG GAGGAGCGG GAACTTAGG AGCGCTGATT AAGATGTCCT ATCTACTGCT 300

5 CTGTGAGGT GCGTGGGCA TGCAAATGTG GGTGAGCTTC GTCTGAGGCT TTCTGCTTTT 360  
TCCAGAGCT TCCCGAAGT ACCTTCCGAC TAGTCCAGAG CAAACTCTTC CCGTTCTACT 420  
TCCACATCTG CATGGGCTGT GCGTTCTATC ACCTCTGCAAT CTTGCGTTCA CAGATGCTT 480  
GGCTGAGCT CACATTTCTG GAGGCGAGCC AGCTTTTACT GCTGTCTCTG AGCGTTAGCC 540  
TGGCACTGT CAGCGGCTCG TGGCTGAGC CCGCGACAC AGCTGCCATG TGGCGGCTGC 600  
AAGCGTGGG AAGAGAGGG AGCGCTGGGT GGGAGGTAC CAGGAGCCA ACAGGTTCCC 660  
GATCTTAAC GCGAGTGGG AGAGAGGAC CCGAACTTCA GTCTCTCTCG CCGAATTTTC 720  
TTGCGCTAGC ATGGGCTGTC GTCTCTTTC ATCTGGGCT GCGTCTCTAG CAATGCGCTC 780  
TGTCTGGTG GCGTGGCTT GGAATATAGG AGCGCTTAGC ATGGGCGCTG CATGCTATA 840  
AATCTCTCT CAGAAAAAA AAAAAAAAAA A 871

## (2) INFORMATION FOR SEQ ID NO: 105:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 404 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 105:

35 GCGACAGTT ATAGCATGCC ATTCTACTT TGTGTTTAT TTCTCATGAC TTTTGTGAT 60  
TTAGAACAA ACAGTGCAC CTTAGAGCTT TGTGCCATG AATTTTCCA TCTCTCCA 120  
AACTGCTTG AGTTACTAG AACTTCACC TCCCATGCA CTGAAGGCA TCTTGTGCA 180  
AGATACAGA ATGGGTACA CATTTAAGT GCGAAAGAT GAAGAACTCT TAATGTTTC 240  
TTTTTAATA GAATGAGCC CCACTTTGGG GACTAATAT GTGCTATGCG CGAAGAGCAG 300  
TCTAAATTT ATTTTATTA AAGAGAAC TCCCCATTA TTTGCTGGG GTTGTGTTT 360  
AATTGTAT ATGAAATTT TTTTGGGCT TTTTGGGCGC ATGG 404

## (2) INFORMATION FOR SEQ ID NO: 106:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1542 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 106:

60.

5 GTGAGAGAG TGGAGCCGC GGGGAGGAG TGTGAMGAG CGAGGCTCA GAGAGGAG 60  
GGCTCTTCA GAGGAGAG GAGAGCCCTG GAGAGGAG GCTGAGAGT CTTAGCCAG 120  
GATGAGGCT GTTGTACT TGTAGAGAG GTTGTAGAG CAGGAGAT CCGGATCA 180  
GGCTACCT CTAATGGGT CCCCTTCT AATGACTTC ATTCTCTCA CTAATGTA 240  
CTTCCTTC TCACTGGC CTGCAATAT GGTATATG AAGCCCTTC AGCTCCCTGG 300  
CTTCATGAT GTCTACACT TCTACTGAT GGCATCTTC CTCTACATG TCTATGAT 360  
CTCATCTG GCTCTCTCA GACATATAC CTGGGCTCT GACCTCTG ACTATGCA 420  
CAGCCCTGAG GCACTAGCA TGTTCGGT GGCCTGGCT TTCTCTCTT CCAATGAT 480  
TGACCTGAT GACATGTA TCTTATCT CCAAGAAA GAGGGGAG TGACTTCTT 540  
ACATCTCTT CATCACTG TCTTCTCT GAGCTGCTG TGGGGTAA AGATTCCTC 600  
GGAGAGAT GCTCTCTTC ATGCAATAT AAATCTCTT GTGATCTA TATATACT 660  
GTACTAGCA TATCTGCT TTGGGCTCT GGCACACC TACTTCTT GAAAAACA 720  
CATGACGC ATCTCTCA TCAATTTT CTTGCTCTA CTGCAATCT CCAATGTA 780  
CTTATCTTC AGCTTAAT ACATATCC AGCTATAT CACTATCT GATATATG 840  
CACTCTTC TTCTCTCT TCTCAAT CTGCTATCA TCTTATCA AGGGAGAG 900  
GCTGCCCC GCACTTAC AAAATGAG TCCAGTAT GCAAGCTCA AGGCCACTG 960  
AGAGATG CTTATATG GCGCACTA AGTCTCTAG GACTCACT TAGGGCTG 1020  
TCTCTCATG CCTCTCAC CTACATCT GACCAAGCT TATGTCTCA GCACTAGCA 1080  
GGGACTGG CCTCCCTCC CCACTGTC TCTACAGGA CAGGGCTT GTTCTCTAC 1140  
CACTTCCC GGGGAGTC CAGGATGTC GCTCTATTC TCTCTCCAC TCCAGACTG 1200  
GGGCTAAA GGCCTTACA GTTATTTCC CTTCTCTGC TTAAACTTG GAGAGGAG 1260  
ACTAGGGCT GCGCCAGAA AGGCTCTCT GGCCTTTTC CTGACAGA AGAGTACG 1320  
AATATCTCA CTGTGACC AGTCTACT CTCCAGCCA CACTGAG CAGTACTTC 1380  
TGGCCAAAG CTCAGGTC GCGGGGCTT GGGATAG CTTGTGAG CTGCTACTC 1440  
AACTCTCT TTAATAAA GTGAGAG AAACAAA AAAAAAAA AAAAATCTCA 1500  
GGGGGGCCG TACCAATC GCGGTATCA TCTTAACA TC 1542

(2) INFORMATION FOR SEQ ID NO: 107:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2127 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 107:

5 GTTACTCAN TCCATGAAA TAGTCTTACT GCAACAAAG CCGTTTATCA AGATATATTA 60  
ACTCTTCTT TTCTTTTTC GAGAGTCTT TTCTTCTGA TGGAGCAAT TCACTGAGC 120  
AAGCAACACA GTATCTTAG CAGAGATCG GCACTTGAG CCAATTTCCG GAGGCAAT 180  
ACATATCT GCACTCTGA GTTGAAGAA TATGACTCC ACTCTTCACT ATATTCACAR 240  
CAATTCAGAC TTGACCAACA ATGACATTTT TAGCCCTGAT GAGCAAGCA GAACTAAT 300  
ACAGATCTT GTACTTACG CTTTGTAGA TCAATATTA TCTATGACT ACCCTTCTG 360  
TCCACAGAG CTTGACACT AAATCTTAA GCACTTCA TATAGCTTC CTGCTCTGC 420  
CTTGCACAT GAGAGAGCA ATGCACTG CTTGAGAGAG ACATATGGA CTCTGGCTC 480  
AGCATCTAG TCCAAATTC GAGCAACTC TAGCATCTC CATCCAGRG CTTGCAGTTA 540  
TTCTTGAGA TCAATGACA GCTGAGATC TGTTCATAT TTCTTATGA TTCTTAAAG 600  
ACTTCATGA AGTCAAGATA GTTCTCTTA AACACTGCA TCAATTTCTG AGCTTCTC 660  
ATATGAGAA AAGAGAGAA TATCTTATC AACTTACGA GTTTTCTG ACAGATATTA 720  
GTAGAAATTC CCGTTTCA GCTCACTCG CTGAACTCT GATTTTACTT CTAGATAT 780  
ATATCTCAG AGATTTTAT GACTTTTAC GTCCATTC TCTGATCT TGTGACAGA 840  
AACTTTCTT TTTCTCTTC ATTTCTTCA AGTGTCTAG CAGATCTG AAGAGCTC 900  
AGCGGAC ACCACAGC TTGAGATCG ACCCTATCA TGAATCTG GAGACTTTC 960  
GAGATCTC CAATGCTCT GTTGGCAG CTTTCTCTT TGTCTGCCAG ACTCTCATG 1020  
AGCATCTCT CTTCTCCAG GACATTTTC CTGTCTCTT CATCCGCA CTCTTACT 1080  
TAGCAATGA CAGGTTCTT AGCTGCGAG TCTGCTTC AAGACATTA AGACAACTC 1140  
TACTAGAAA AGACTTTTC TTGCTCTG CCACTTCA CAGAGGCT GTGAGAGCA 1200  
CAATCTGC TCTTCAATG GACCTTACA CCAATTTCA GTTATTTCA AGCATCCAC 1260  
CTGCAATTC CAATTTCT GAGATCTCA TGAAGAGC GTCTCAAC TACTAGAGG 1320  
CTTGAATCT GATCTCTTC CTGCTTCTT GAGAGCGAG GTTCACTGG CATTCGAC 1380  
CAATGTACC TGGATAGCT TTGCGGGAG GAGAGACTT CTTCTCTTC GACTTCAAT 1440  
GAGGTGCA GTTGCCTACA CCAATTACA CCAATTTCA GATCTCAG AGATTAAT 1500  
AAACATAT ATCTTATCT GACTTAAAG KRNAAKMM KCTCAKNSIA TTATATTSW 1560  
CMBBARGSH WYAAWSTK SAGCTCTWCC KSRSTWIKG WBCCTCTAGA ATTTGRLAK 1620  
CMTTUSCT RMGGAUKS GGLSAGGCC AGAGACTGC ATTTCTTCT CTTGTTTTH 1680

TTTAAATC GCAATATGA ATTCTTACAG CTTGAGGCA GAGCTGTGCC CAGATGTGAA 1740  
AGAGACTTC AGTATCAGCC CTACTCTTC TCTCCAGGA AGGACTTCTT GGCCTCTGTG 1800  
GCACTCTTC CAGCCAGCC CTGTGTGTGA ATCTTTTGTG ACCTGTGCA ATGCCAAGG 1860  
AGGGTTTTT ACATCTCTTA AGGACCTGA TCCGACGCA AGTACGATG ACTTAACCTC 1920  
TTAAGCCAG CATATCTGT TACACATTTA CAGATGCTT CTTGAGTGTG TGTGTCTGAT 1980  
TTTTCTATCC TGTGTATGAC CTGAGGAAA TTATTTAGAC GTATATGTA TGTCTGTGT 2040  
TTTAACTTG ATCATGATCA GCTCTGAGT GCACTTCTT CACTACTGT ACATACCTGT 2100  
GACCACTCTT GCGAGTCTG CAGTCTTTA TCATCTGTT TAACTCTTG TGGCAGAGT 2160  
TCTCTTCTCC AAATAAAATT TATTATAAG ATCATATAG AGATATAT ACATTTTCA 2220  
TTGTTTCTTA GATCTCTACC ATTAATGCA ATTGTGACC TGTAAAAAA AAAAAA 2280  
ACTCGAGGG GCGCCGATAC CCAATGCCC GATATGATCT AATCATC 2327

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1062 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 108:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 108:

60 GCGCCGAG GCGACAGC GGTTCGTCA GCTCTGGTC CAACCGACT AGCAATATC  
120 TTCTCATCC TCATCATGT CTCTCATC CGATCTCG TCCAGTCCC TCCTCCGCC  
180 ACACAGAG TCGCAGAGT CCACTGTAG TTCTCTGGA GTTCTCGAA GATGCTCTC  
240 CTCTTCTCG TCATCATCTT CTCTTCTG TTCTCATCC TCATCATCA GTTCTCGAG  
300 CCGCTCAGA ATCCCATCC CCGCCGCGA GAAATGACA GAGGCGCGG GTACAGCTCT  
360 TATCTTCAC ATGACATTA CCAAGGCA AGATGCTAC AAGAGAGG TGCATGAA  
420 GAAGAGGG TGGTCTTAT TGAAGATA CTTGCGCCA TCACTCGATC AGAGTGAA  
480 CAGAGTCTT CGTTTTCG AGAGTTAG GATGTCACA TCCACTTCG TGTCCAGGG  
540 GAAATACG GCTTGTGAC TTATCTCAT CTCTGAGAG CATTTCGAC CATGAGAT  
600 GCGCAGCC TCGCGAGCG AGATGAGG CCGTTTATC TCTGCTTGG GCGCGAGG  
660 SHGTGCTCA AGAGAGCTA TTCTGATCTT GACTCGACC GCGAAGACTT TGACCCAGA  
720 CCTGTAAAG CCAATTTCA TTCTCTTAC TTGACAGT TGTGAAACA GCGCGAGAG

AACTCAGA GGTAACTTG GGCCTTCC TGTATCTTT TTCTCTCTT GAGGTGCCC 780  
AACTCTCTC ACCCTTCC CTTACTTAG GCGAGAGC TGTACTTAG ATACTCTTT 840  
TATTAAGAA TGGAAAAG TGAATATAA ATATCTTGA ATCAGATTTT TTAAGAGGG 900  
TATTTCTTT TTTATACAG GTATGAAAC AAGTTAACTT CCACTCTTAT GTAAATAGG 960  
AGGCTGAG GGAATCCCA GTTTTGGAA CATACTCAC TATCGAGCT AATAACATC 1020  
AACTAGAG NAAAAAAA AAAAAAAA ATTAAAAA CT 1062

(2) INFORMATION FOR SEQ ID NO: 109:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2539 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 109:

60 GAGAGCTCA CACTCTTTT CCAATATCAC TCAGATGTA GTGACATAG CAGGGAAGA  
120 GCACTACTT GTTTGTGTA GTTTGTGTA TCAATCTAT AACTAAGAG AGGATTTAT  
180 AGGCTCTCT CTTATGAGG CCAATCAGA AATTGTGCT GTCAATTTT ACATATGAT  
240 AACTGAGAG TGGGATTA ATATGACTA TTGTCTGCG CAGGCTTACA TTGCTCTAG  
300 TGAATTTCT TCAAAATGA AAGTTTCTC TTCTAGACTT TTAAGAGGA TACCCCAAC  
360 YAACTGAA ANAKCTGSM CTCTCHSTOC SKTRRKRVC GTATATATC WAKKSAKVM  
420 CCGTGTGTC RGTATATSK TCAATAGGG AACATTTGAG GAATTTTCT CTTTCTTCA  
480 TCAATCACA CACTCTCTT TAGAATTTGA CAATTAATT TCTTCTCTT TTCAGAACAG  
540 TAAAGAAAG GGTAAAGAC TGAAGGAAT CTCCATCTT CAGTGCACG CAGGCAATGA  
600 TCGTTTGA AATTATGCG AACTCTTCA AGCACTCTT TTAATTTAG ATGATATAA  
660 TATGACACA ATATATGAT GAAATACTA TATAGCTGCG CAGCAATTT TACTCTGAT  
720 GCACTGTAG AATTGATTT CATTGTACT ATTGTGTC TTAAGATAT CCACTCTTT  
780 AAGAGACCT TTGGAAAA CTTGAGGGG CAACCTCTG ATCTCTCTT TCGGAGKST  
840 ARATTTTGA ATATATVACT RMTKSNVTV AVGTGTGA CNGEIMANTA TTAATATTTA  
900 TACATTAAT YMTTATTSK KPAKATAT CCAAAATAT GAAATATAA CTTATATTC  
960 AATTAATCT CCGTGAAA TTCCGAGAG CTCACAGGG TAACTGGA TCTGAGTAA  
1020 CTTCTGAG TTAATATAA GAACCTTAA GTTCCGAC AGTGGAGAC AATTATCAG  
1080 AACTTAAGA TATATCTCA GACAGCAC TCAAGCTCT TAAATGCTTA TCTCTGTAC

CTCTAGTAT GGGGACATC AATTCAATA GTTCGAGGA ACACCTGCT GACATGATA 1140  
GAACTGACTT ACCCAATCT GACACCTCTT CAGCTGAGT TCAITTTTGG AGAATCAAT 1200  
GGAAACAGG GGGGAAGAT ATGAGCTTC GTTCACCAAT CTATGAGCC CTCCACCTGC 1260  
CTGACATCA GTTTTCTCT AATGTATAG CATCTCTGA GTCTCTGTAT ATCTCTCTG 1320  
TGTATAGGT TGTATATAG CGGTATGAAA ATGAGCGAAA GGCCTTAA GCAATATGGA 1380  
GCAACATTT GACAGACCA AGTCAAGTA ACTTGACTTT GTTAAACATA AATTTTGATA 1440  
TAAACACGA CTCTGATTTA ATGTGAGCA CATTATATTA ACTCTATACR AATATGTAG 1500  
HCTTCTAC AKATATATCH SHAKITOTGG AATWSSSTA KGSWGWCK TAWBERTWG 1560  
GMMPTTTHK RHTVGAHWH YGGMHOCAG AAAAGCCCT AGGTGTATG TAGACCACTT 1620  
AATCACTAAA TATCTTTTCC TATAGACTC CATATGATAC ATTAGCCAT GATAATCTAC 1680  
CTTTTAAAT GGGCCCTGTT TGAATCTCA AGCTTTGAG ACTTAACTCT TCTTCAGAA 1740  
GAGAACCTTG AAGTCCCAT GTTCTCTTTT GGTGTATCTC TTTTGAATGC ACTCTGANT 1800  
TGTCTTCACT TAAKATCAT TTAGCATAG CATTATATAT CACTGTGAT CTCTACTTGT 1860  
TGGTGTAT GAAITCTTTG AAGATATAT TTTGAAGGG TGTGGAGGA AGGAATCAT 1920  
TTTATTAAT GTTCTATGA AGCCACAT TGACTTGA CTATATAGAG TTTTATGAT 1980  
GTTAAATAT TATACTGAC AGTTACAGA AATTACCGA GAAAAGCTTG TGACTCAC 2040  
AATCAGANT TTCACTGAG ATTTTCTCT TCTTCACTT AAGAAACAA ATGCACAGT 2100  
TTGATGAAA AAGCTGCTG TTGTTCACA TCTCTTCTCT GTTTTACATC CTTTGTGAG 2160  
CTTACTCTT CTATKCTTT TTAGCAGTA TTGTGTGAC ACTTCTGTT CATGCTGAG 2220  
ACGATATAG AGCCATGGA TACTGACAC TGAATTTCT GTTTTCTTC TCTCTCTTT 2280  
TCCATGACT TTATATACT CTCTACTTG ATTATAGC AAAACTGGA AATCTTCAA 2340  
AATATGTT GTGTATATC TAGAAAAATA TGGAAATAT TGTCTTTTAT TTGTGTGAG 2400  
AATATCAT TTGTATGTT TATTCATC TAAATTAAT GTGAATTTG TTAAATTA 2460  
AATWGSAC AABTBGHH GGGTCCAAA CHTWTCGH KAAITCTCT WAARATYK 2520  
ATAACASCT TCAKATC 2539

(2) INFORMATION FOR SEQ ID NO: 110:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1751 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 110:

5 AGCATGAGC GATATGCCCT GTTGGCCAGT ACCGTCTGG GCTGTGTCCA AATATGCTT 60  
GGTTTGGCG GATCTCTGCT GTTGTCTTAC TACGTATTTG CATTATTTGG GATCAACTTG 120  
TTTAAAGGCG TCAATTTGGC TCTTCTTGA AACAAGGCC TGGCCCTTGC CATGTGCTCG 180  
GGGCTCTTG GAGCTTGA GAGCTGGAG CTACTGGGCA ACMACTTGA TACTTTTGG 240  
GCTTGGCTTG TCACTCTTG GAACTTGATG GTGTGAGCA ACTGGCAGT GTTTCTGAT 300  
GCAATGCGC GCTACTGAGC CCGTGTCTCC AGATCTAAT TTGTATTTTG GTGCTGTG 360  
TGTCTGTGA TCTGGTTCAA CTGTTTCTG GCTGTGATTC TGGAAACTT CTTTCAAG 420  
TGGAGCCCG GAGCCAGCT GAGGCCCTT GCTGGAGCC CAGAGGCCAC CTACAGATG 480  
ACTGTGAGC TCTTTTTCAG GATATTTCT GAGAGCCCG GAGAGGATGA GCTCAAGAG 540  
AGGCTAGCC AGCACCCGCA CTGTGTGCTG TCGAGGTGAC GTTGGGCTG CATTCCAGC 600  
AGGGCGGCA GAGAGAGAG GCTGGCTTAA CACAGTCCC CATCATGGA GAGCGGCGCA 660  
TGTGTGCC AGCCAGCAG GAAGAGCTT TTCTCTGAC GAGCACTTAA GCTGGGAGCA 720  
GGAACAACT CTTTGTGCTG TGGCCACAA ACCATCTACA GAAAGCTGC TGTGTCTTCA 780  
GGAGCGGCG GTGCCCTCG CTTTCTTTTA TACTGCTTC AGTGAATTT CCTGTGTGA 840  
CTCAAGAGG ACTTTTCAGA CAATAATGA AGAAGCAGC GCTTCCCTG TCTCTGTGAG 900  
CTTGTGCTG GCTTTTCTG CCGCAGCC TTGGGAGCA CAGGCTGAC CAGGCGCTGC 960  
ACAGGTGAC GTTATGCTG TCTATCTAT TCAAGCTG GATGATACT AATACCTCG 1020  
ATTTAGCCC AGCACACAG GGTAGCTTC AGTTTCTC TCTTTCATA GCTGTAGGC 1080  
CCTTCTGCG AATGCTTCT ATCTCTTAA ATCTATAT GTGTAGTTT TCTGTGAT 1140  
CCCAAGCTC CATTACTGC CAGCACTCC CCAAGAGAT GCGCTGCTCA TCGAGCTGG 1200  
GCTTTGACT CCAACTGCTG TACCCCTCTT GTTGGAGCC CTTGTGCA AACTTTCAG 1260  
CAACAGCTT TCCAAATGA AGTTGTACT GTACAGCTT TACATCAGC AACAACAAA 1320  
TCTATGCT GTTGAAGGCT CTGCTCATT AAGATGCAAT AATATGTA GTACATAAA 1380  
ACAGCATAG AAGAACTTA ATGCTTAT CTCAATATG ATGTCTTACT AGAAAGCCA 1440  
AATATATTA GATATGTAAG AATTCACCA GCACTTTGG AGCCCGAGC GGTGTGATCA 1500  
TGAAGTCAG AGATCAGAC CATCTGCT ACAGGTTGA AACCCCTCT CTACTAAAA 1560  
TACAAAAAT TGGCCCGGCG CAGTGGCGG CCGCTGTGTT CCAAGCTACT GGGAGGCTG 1620  
AGGCGAGGA ATGGCTTGA CCGCGAGC GAGCTTGA GTAGGCCAG ATTGCCAC 1680

TGCAGTCCG AGTCAGCCT GCGCGACGAG GCGAGACTCC GTCTCAAAA AAAAAAAAAA 1740

AAAAAAAAA A 1751

5

(2) INFORMATION FOR SEQ ID NO: 111:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1117 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 111:

AATGTTGCG TGGTAGCAT TGGGTAAATT CTAATATAG AGTCTCTCG AGACAAATGT 60

CGTAATCTA ATCCCAACA AATTGCTTT TTATATGTT TACTGACAG CAGAGAAATT 120

CAAGAGCCT TCGTAATATC TGTAAAGCG CAGTACAGC AGGCATCAT GCGTGGGTGT 180

ATGGGGAAAT ACCAGCTTTT ATTCACTCTA ACGACATTA CATTGACAG ACCCGGCGAG 240

AAATTATCA TAAAGCGTTT GATCTGTGC AATCTGCACA TGTCTCTGCC ACAGAGGCT 300

TCAATCTGTA TGGCTGGCGC TGGGTGGA GAACTGCAAT GTTTGTGACT AATATCACA 360

CAGTGAACAC TAGTCTGAAT GTATACGAA ATAAAGATGC CTTAAGCCAT TTTGTAAATTG 420

CAGAGCTGT CAGCGAGT CTTTTTNGA TAAAGTAGG CTTGCTGCGC CTGCTGGCTG 480

GTGGCAAT TGGAGCTTG CTGGGAGTC CTGTAGAGG CTTCTGTGAT GCAATTTGAGA 540

AGTACTCTCG TGAAGCTGTT CAGGAAAGA AACAGAGGA TCGAAGGCA CTCATGAGC 600

TAAAACTGGA AGAGTGGAAA GCGAGACTAC AAGTTACTGA GCACCTCCCT GAGAAATTTG 660

AAGTAGTTT ACAGGAGAT GAUCTTGGA ATGATGCTTA GAAATTTGAA GCACCTGCTAA 720

ACCTTCTTAG AAACCTTTCA GTAAATGATA AACAGACAA GAACTGAAG TCGTCTGAAAC 780

TTGAACTGCA CTGGAGGCT GAAGGAGCT GCGATGTCG ACGATATCCA ACAGCAGGCG 840

CACTCTTTGG TCAGCTGCT GACAAATTTA AGTCTGCTTA CTTGTGCTGG CAGTGGCTTG 900

CTCTGTGCT TTTCTTTTCT TTTTACTAA GAATGGGCT GTTGTACTCT CACTTACTT 960

ATCCTTAAT TTAATACAT ACTATGTT GTATTAATCT ATCAATATAT GCATACATGA 1020

ATATATCCAC CCACTGAT TTTAGCAAT AAATTAACA TTTGCGAAA GATTAAGTT 1080

GAATTTTACA GTTAAAAA AAAAAAAAAA AAAAAA 1117

55

(2) INFORMATION FOR SEQ ID NO: 112:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1117 base pairs

(B) TYPE: nucleic acid

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1313 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 112:

GCGAGGTT TCTTATATT TTATGTAAT TTAAATGCG TATCAAAATA TTTATCTTG 60

TTTGAGACTA CCAACATAAC TACATGTTGA AGATGCTTCA CAGCAATAT ATTGCTTTA 120

ATGTCAATA ATTTCACCA ATGTGCTTA CTTTATAAA GTATTAATTT TGTAAATAT 180

TCAATTAAT AGTTGTAAC CTTTCTAT TTATTAATA CTTAATCAT GTTACTTTT 240

TTTGAAGA TCGAGCAAT CTTTGAAC ATGAATGCG GGTCTCAT ATGCTCAT 300

TACTAGAAA AGAAGAAA GAATGCTTC TTATGTAAT TGGCCATGT TGTAAAGGCT 360

TGTGCTCAG ACGGATCTT GCAATTTGAA AGCACTTGT TAAATATGAG AAGGCGACCA 420

TAAACAGC TTACAGAGT GCTCCATGT TACACATGA ATTACTTGA TTGATGCTTC 480

GATTATTTA GCGAGAGCT ACTTGCATG CTCAGACCA TGAAGAGCT GAAGATATA 540

AACTTACA GTCAATATC CACAGAGAC TGTGTGTGC TTAAAGCCA AGTTACATA 600

TGTATTTTT GGTACTGCT TCTTCAGA GTCAATATC TTTTCAAG TTCTTGTGT 660

TCAAGAGT TAGTGACAAA GCGAGAAAG ATTATCAGC CATCTAATA GAGTGAAGA 720

TTTGTATCT TAGAGACT AGTTTGGC AACTTAGAT TTACGTTAA TTTTACATA 780

GTATTTGCA CTCATGAAA ATATGTGAA AACATCTGA TTATGTAAT TATCTGCGC 840

CTTTGTATA AACTGAAGT TTTGAAAT GGTGTGCT GCTCTTCAG CCTATGATA 900

TTTTTGTAA ATGGAACCAT GAAATATGT CTGATCATC CATACAGAC CAAACATTTT 960

AATCAAAAC AATGCTTCA TCAAGTAAT TGTCACTAT GCGTGGTAT GTTTATAGC GAATTTGAGC 1020

GACCACTGA AAGGAAATG TGTCTAGCT GCGTGGTAT GTTTATAGC GAATTTGAGC 1080

AAGAGAGC CAATATGCT TTTCTCTTTT GAAGTTTGT TAAATATAT TTAATGCTC 1140

TTTTTTTAA TTATATATG TGCATGTTA CAAATATGT TGGATGCT TTTGACCTA 1200

AATGCTTTT TTGTTATAG AGATGTGTA CTATTTTAT TTTTATAAA TGTATCTCC 1260

CTTTTAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAA 1313

55

(2) INFORMATION FOR SEQ ID NO: 113:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1654 base pairs

(B) TYPE: nucleic acid

60



336

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 113:

```
5  ACGGACAG AATATCTCTT TTCTCTCTT CAGTACAG AGGCTTTCT CTACATTTG 60
   CGTCTACCT TTAATTTAA AGCTATCTT TTCTAGTAGT ATTATATAT GCAATGCGA 120
10  TATATACAC AACAGCTTT CATTTACAG TGAAGGAG CATCTCTTA CTTAATATG 180
   TTCTCTACT TTCTCTCTA TTATAGGAA ATTATACAGA TTCTAATAT ACCTATATT 240
15  TTCTTTATT TTATTTATC CAGTACAGA ATCTCTTTT CATCTTTTA GAATTTCTT 300
   CATAGATCA CAATATCTA CATTTCTTA GATGATATT TTCTAGAT TTCTAGATG 360
   CAGTGGCAT CAGTGGCTG GTTGAAGCTA GAATTTTTC TGGCTCTCT GACTGGTAA 420
20  GCTCTCTCT CGGACCTG TCACTGGTG AGGAGATGA CAGTGGTCA GATGGAAG 480
   AGAATATAT GAATCTCTT GGGCTCTCTG GTTCTCTCTT TGGCCAGAG TTATTGAAA 540
   GCAGCGGAA TCACTACT TCACTCTCT AGCTTTCTA GCTTTTCTT TATTTTCTG 600
25  TCTTTAGAT TCTCTCTCT AAAGAGGCA AGAATTTAA CTCTCTCTT CTACAGAC 660
   ATGGCATCT ACCGATGAG TCTTTATCT GGTGTTAGG CATATAGAT GCTGATCTC 720
30  AGTCTCAGC CAGATTTTC TTCACTATC GTCCAGCTG AGTGTCTCT TCTCTCTG 780
   CCAATTTCT TGAACCTCT GCTCTAGCT TGGCGAGGG AGAGTCTAT TTGCTTTCT 840
35  TCTCTCTCT TCTTAGAAA AGCACTTTT AATATATCT TTCACTACT TGGGGTCT 900
   TTGTGATTT TTTTCTCTT CCGAGACT CCTGTTCTT ATTGGATTT GTATTTTAT 960
   ACAAATATT GAATTTTAA AGCTCTTCA CATATTTAA TTATGTGAA AGGAAGAAA 1020
40  GAATCAGAA AAAATATTT AATATCACC TCACTTACA AGTGTCTAG ATTATCTAT 1080
   TGGGATCTT CTTTCTTAA GTTTTTCAG ACAACCTAG ACTTAACTG TGTCAAGAC 1140
45  TTCTGAATCT TTAGCAGTG CTATATATT CTTCTAATG ATTCTGTAT TACTTCTTA 1200
   TTCTTTATC CTCTCTCTC TGAATTTAA TGAATTTGA ATTTAGCTG GATTAATTA 1260
   AAAAGTAGT GTATAGTAT AGTATCTTA GTTCAGATCA AGTGTGTTA TATATCTTA 1320
50  TTCAATATA TGAAGTTAG TATTTACTA GGTTTACTA AATTTACTA ATATCTCTT 1380
   GAATCTACT TGTCTCTCG CTAGAAAAA TTATTAACAG GACTTTTAG TTTGCGAAC 1440
55  CAATATGTA ATATCTCTC TTCTAATAT TGGGCTATC AATATATTT AAGAATATG 1500
   GATTTTATT CCCGATAT GGTGTCTCT TTATGATTT ACCGAGATG ATGTATTAG 1560
   TAAATAGAT TTGTAAATA TGTAAATAT TCAATATAA AATATCTTT GACTTTTTC 1620
60  CAAAAAAA AAAAATAA AATGAGGG GGGC 1654
```

337

(2) INFORMATION FOR SEQ ID NO: 114:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1171 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 114:

```
15  GCGAACTTT CCGCAAGGC TTGCAACTT GCAAGCGGA ACCTTGATC GTTAAAGTT 60
   GGGTTGCGC GGGGCTCTG CCGCAAGAG GCGAATGCG GTTCCGAAA CGTTGGCTT 120
   CAACTGCTG GCAAGCGGC ATGCTCTCA CCGAGGAGG GGGCTCTGG CTACAGGAC 180
20  CTGAGCTCA TCTCTCTCG CCGACTCTG GGGGTTAG GGAATTTTA GACTGTGAG 240
   GAGTCTCTG TGGATGCTT GTTGAATTC TTACCGAGG GGTGAGCAA AGAGAGATC 300
25  AACACTCTA CCGTCAGGA AGCTTATCT CAGAAATCT TTAAATCTG CAATGACTT 360
   GACCATGGA GTCTTATAT CTTCTCAAC ACAAATGCA CTGAGATTT TCACTGACA 420
30  GTGATTTCC TCGGAGGCA GTTTGAATC AGTATAGAT CTTTCAAT CAATTTAGAC 480
   TCTCTCTCG TCTTTATGA ATGTTCAAG AACCAATGA CTGAGATTT TCACTGACA 540
   ATATCTGGG AGAGCTCTA TGGGATTTT CAGGAAGCT TTGATGACT TTGTAAAG 600
35  ATCAATGCA CAGGAAGCC AGAGGAATC CGAGGGGAG GCTTCTTAA GTACTGCAAC 660
   CTCTTGTGA GGGCTTTAG GCGGCTCTT GATCAATCA AGACTTTCA AGGTATATG 720
   TGTTCAGGT TTTTCAATGA CTTCTCAGC ATTCAGAGC AGCAGGAAA ACTGAGTCC 780
40  TATTTGAGA ACCATTTCT GGAATTTGA AGAGCGGAG TATGATATC TCACTGACT 840
   TCACTAGTG GTAAATGGA GCACATCTG CCGATGCGA CATGAAGAA GACAGACTT 900
45  AAACCTTAT ACATCTCTG CTATCGGGT GTTAGCTGAC CAATATCTA TTCTTAATG 960
   GGTAAATCT ACTGCTATT ACCAGCAGC CCGCTATGTA GCAATGCAA ACTTTAGCA 1020
50  TTACTAAT GCAAGCTTC AGCAGTATT CACTCTCAG CAGAGACTT ACTCCACTG 1080
   GCTACCTCG AATTAAGAT CATTTAAAA TGTCTCTCG GGAAGCAAT TCAAGACAA 1140
   CAGGAGAAA AAAAAAAAA AAAAAAAAA A 1171
```

(2) INFORMATION FOR SEQ ID NO: 115:

(1) SEQUENCE CHARACTERISTICS:

338

339

- (A) LENGTH: 842 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 115:

5 GGTCTGCGCC GGAATGCAAT GAGCTGCGCA TGTGTGCTT AGTGAATGCG GTTTCGTGCG 60  
10 CTCTCCGCGC TTTCCTCGGC TGGGTATTTG CTTGCGACCA TGGGCGCCGA GGGCAAAATG 120  
GGCAGCAGAG GGAAGAGCA GATATTTGAA GAGACAGAG AGACTCTGAA GTTCTACTCG 180  
CGATCTATAC TGGGGGCGAA TGCATTTAC TGCCTTTGTA GGTTCGTCTT GTTTTACTCA 240  
TCTGACTCAT TTTCGCGCTG GTTGGCGCTG GGTCTTATGC TGGCAGTGTG TGGGGCGAGC 300  
TACCACCTTA TGAAGTCAAT GGGAGAGCA GCGTTCTCTG AGGATGCGGC CCGTATGGAAT 360  
20 GGTGCAATCG ACCTCAACAT GAGCAGGCG ATGCGAGAGC ACCTTAAGCA TGTGATCTTA 420  
CTGACAGCA TGTGCAAGT GCTGAGTGC TTCTCTCTCT ATGTCTGCTC CTTCCTGCTT 480  
CTGGCTGCGC GCGGGGCGCT TTACTCTCTG TGGGTGATG TGGTGGCGCC CTGCTTCACT 540  
GCGAGCAGTG GACCTCGAG ACCAGAGAC AATGAGAAC GGCAGCGCGG ACGAGAGCGG 600  
GCGCAGATCA AGCGTTATA GCGATTGCA TTGTGGCGAC AGGCACTGCG CCGTGGGTGG 660  
30 CTCTCTGCGC GTGCGAGCGC CTTGATGCTT GAGCAATGA GGGTCTATGC CAGGGGCGCA 720  
AAGCACTGCT AGGTATTTGG TATACTTATA CTCTATAGGG TCGTGTGATA AATGCTTTAG 780  
35 AATGTGAAA AAAAAAAAAA AAAAACTCG AGGGGGGGCG GGTACCGCAAT TTGCTTANA 840  
AT 842

(2) INFORMATION FOR SEQ ID NO: 116:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1640 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 116:

50 GGCAGAGCG GCGCGCAGCG GTGCGCGCGG CCGCCCGCGG CCGGAGCGCT TCCCTTTCCG 60  
GTGCGGAGCG GCGGGGTGCG GCGCCAGCGG ACCCGCGCGC ACGGAGAGCG GGAAGAGGAT 120  
GATTTGCGCG GCGCTCGCGC CCGGATGAAA GAAAGAGGAA GTGATCGGAA AATCTGGGCT 180  
AATCTGCG AGAGCGGATG TCTACTACTT CAGTCCAGT GGTAGAGGAT TCGAGAGCAA 240  
60 GCTCATGTT GCAAGGTACC TGGGAATATG TGTGATCTC AGCAGTTTTC ACTTCAGAAC 300

360 TGGAAAGATG ATGCTTAGTA AATTACAGAA GAACAAAGAG AGACTCGAA AGATCTCTCT  
420 CAATCAAAAT AAGGTAAAC CAGCTGTGAA ATACAACTAT GCGATTAGA CAACAGGCT  
480 CAATTTTCAA ACAAAGCTTA ACCCAAGTTC ACAATCATC CTAGTAAATA AGTAAATCA  
540 GACCCACAC GATGATATCA ACACCAAGT CAGCTTTTCT GGGAGAGAG GCTACAGAGA  
600 CTTTAGTCCA TCAGTGTAAA CAGAACAAAT TATATAAAAC ATGGAACTAC CCAAGGTCT  
660 TCAAGAGATT GGTCCAGTAG CATTGATGAG AGCCTTTTAT CTGTGTGTC CAGTCTTTG  
720 CACAGAGCT CTGGCGAAT CACAGGCGAA GTCTCGCTG CTGTGGAAA GAACTGCTG  
780 TTGTGCTTAA CAGATCTGAA CCGCTCTGCA AGCTTTTAT TGTGCAAGT GAAGACTGAG  
840 GAAACAGAG AGCGATGACA GCAATGAGCG AGAATATGCG AAGAGCACT GATGCGACAC  
900 ATCTGTGCGC GAGCTGCTGA TACAGAGAG ATGATATG AATGAGACAG TGGAGATGAA  
960 GCTTAAGAT ATGATCAAGT AACTTTGCGC CCACTTTCCG CAGAGAAAA TTCTTAGAAA  
1020 AATTGACAAA AATGTTTCCA CTGCTTTTG CTTGTAGGAA AAAAAATGTA CCGAGCACA  
1080 TAGAGCTTT TAATAGCACT AACCAATGCC TTATTAGATG TATTTTGTAT GTATATATCT  
1140 ATATTCGAAA AATCAATGTT TATTTGAGT CTTAGGACTT AAATTAGTCT TTATTGTATA  
1200 TCAAGAGCA CCGTAAGATG AGCTGAGCT TTGTATGCGA GGTGCAATCT ACTGGAATG  
1260 TAGCACTTAC GTAAACAAAT TGTTCGCGC AGAGTTTAAA TAGAGACAGA TCAGGAATTC  
1320 TAAATAAAT TCCAGTTTAA AGATATTTGT GACTTCACTG TATATAAACA TATTTTATA  
1380 CTTTATTTGA AGGGGACCG TGTACTTCTT TCGTGTGCTA CTTTAAGAC AATAAATCA  
1440 TTTATTTCCA CAGAAAAAA AAAAAAAAW WISTYGARBR GSRGCKELSN AYMAURWCC  
1500 CCGATWRGGS MTCSTWTKA YTTAGATGCA ACTCTGATCC CCGGCGCTTA GTTTTGACAT  
1560 GGGAGGTGGG AGGAAGATAG CCGTATATTT TCGAGTATGA ACTATTGCTT CTGGGAGCTT  
1620 GTGAGGAAT GTGCTTTCAC CAGATTTCT AGGATTTCT GCGTTTATA TCACTAGCC  
1640 TGTGTAATTT TTTTTCCTCT

(2) INFORMATION FOR SEQ ID NO: 117:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 952 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 117:

60 TGATTTAGN AAACACTTG GAAACTCAT AACCTCATCA GAACTGCTT TTAGCCACAC 60

5      TCTGACCTT CTAGATGAGT AAAAAAAAAA TGAAATATAT TCTTGGAAAT TAAGCAATTT 120  
ATTATATATT GCTATTTTTT TCAATGTTCT AGATATCTTT AAATTTGTAT TTGTGGATTC 180  
ATTATCTCTG CAGATATCTT TATTAATATT ATTGGCTCTA TGAAGACTGA AGTATGTCAG 240  
CTTTTGTGTG AACTTTAGTG GACTCTGTG AGATGTATGT TGTACTTTGT ATCTCTTAAT 300  
CTAAGATAG TTTTATATA CTCCAAAGA AAATCTGCTC TCTTTTCTGA TCTAAATAT 360  
CATCTTTGGG GTAAAGATTT AGTCTCTA AGTTGTCTAC AGTTCTATAG GTCAAGGGA 420  
GTAAGCTTGG CACTGTGACT CTCCCATCC ACGACTGACA GATTTCCACA GAGTGTATTT 480  
TAAATCTCC AGTATGACAT GCTGTGTATG GAGGGGGTAA GGGCTGGGTT ATTAATATAC 540  
AGGCTGCTGT ATTTTACATT GATTTGTGGG GAAGGGGAGC CTGGAGAAA CAATGTACT 600  
ATTCTCTTTT TTGAAGACGG AAAAAAAT ATTTTGTGTT CAGTAAATAT GTTAGAGAT 720  
TCAATGTCTC CTAGCCACA GGAACAGTT CCACTAGAAA GTGACATGTG GGAATCTAAA 780  
ATTTCAGAAA CATTGGGGA AGGAAATAT GCTTTCTCT TAAITGGGAG ATGTGTCAAT 840  
GGGGGGGGGG GGGCTGTGTT TTGTGGGAT GTGTATGTT GTATGTACGC ATATATGAC 900  
GGATCTCTG TGAATTTATA AGTTCCAAA ATATGATATA ATCTTGTGTT TTTGTATAT 952  
30      TATCTCAATA AAGGCACT GBACTCAA AAAAAAAAAA AAAAAAAAAA NN

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1256 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 118:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 118:

45      GAGTCTATAG GTAAAGAGC TCTGTATCG TGGAGGGGC COTGGAGGC TGCTGGGTA 60  
CGGGCTGTG CTAGCCACGG AGAGATGCC TGTATCATC AGCTTGGGCT GAGAGGGGCT 120  
GGGGGGGGGG GGGGGGAGC TGAATATAT GCTGTCTAGG GGGGGAGCG TGTGTGATTC 180  
ATCTTTGTAC CCGCAGATC TACAGATTT GCAATGTAT AGGCACTCAA GAATATGTG 240  
TTGAATATAC GATGGCTGTG AAGACAGAG GACTTTTAT CTTTCTGTAC COTTCCTCT 300  
ATGACAGCC TCTCTCTGAC TGCATCTG ACTCTTTGAG AGCAGAACTC CTCTAGGGA 360  
CTCGATATGG AAGAGAGCT GCGAGAGAC ATCTGTGGTG AAGCAGGGCT AGCTTTTAT 420  
GAAGTACACA AGCTAGGAT GTTGGACCA GAGTTTACC AGCAGACCAT AGAGTGAAG 480

5      GAGAGTGA AGACTTTCT GGAATAAT GCCCACTTTC AGAATATAT TGTGTGTTTA 540  
ATTGAGCTTG TTGATCACT TCGAAGAA GCGAATAATG AAGATGAA GGCATGCT 600  
GCTCGAATCT TGTCTAATC TATAGCAAG CAGAGAGAG CTCACAGCA GCACTTCAA 660  
GCCATATAG CAGAAAGAA AATCGACTA GAAGGTATC GGTGTGAATA TGAAGTTTG 720  
TGTAAAGTAG AAGCAGACA AAATGATTT ATTGACCAT TTATTTTCTA GAATGAACT 780  
GAAATTTTGG CTTTATATGT AGGAAGGCAA ACAAAMAAA AGCTCTCTCA AACCAAAA 840  
ACCTCTATAG CATTCAGCG GCTTGACCA TGACTATGT CACAGAGAT GGGTGTATG 900  
GAATCGAGCC CCTGAAAGAC AGCACTACAA GTCTGGGGGA GCAATTTTAA ACATCATGTC 960  
ACAGCTGCTG CTGTGGGCC TCGAGTATAC GTTCTCACT CTATGCTTAA GTTGAAGTA 1020  
AGCAATTTCT AAATTTTAT CTTTTTTTT GTAAATCTAC AAAGCTTTTG AAGAGAAAC 1080  
AATAAATTTT TGTTTTCAA TGGTTGATG TACTTTTTT CTTTGTGTC TTCAATATG 1140  
TTTAACTCT CATGAGAAA CCTGGATTC TGTATGCTT AGTCAACAA ACAACCAAG 1200  
CATGTGTGAG CAGTACTCT TAAATGTAT CACAGAGTG AGTCAGAG TACAC 1256

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1143 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 119:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 119:

40      GCGCTAGCA GCGGGCTGG TCTCTCTGG AGCGGGGGC CCGAGTGGG GCGCGGCAT 60  
GTACCTTCCA CATTCAGTAT TCGAAGAA GTGATCTGAA CTCGTGACCAT TCTTATGCA 120  
TACATTAAT CAAATATATG AGTCTGACTA CTTGACACAC TGCTCTGAGC AAACATGAC 180  
GTTCGAGTTG CCGACATGA AGTGAATCA ATATACCGTG TCAATGACAG CCGGGTATG 240  
TGCTGACAT ATCATTTGG AGTTGGCTTG CTTATATATG TCTTACTCAG CATTCCTTC 300  
TTCAATGTTT CTTTGTCTTG GACTTTTACA ATATATATAC ATATCTGGG GATGTATATA 360  
TTTTTGCAT CAGTGAAGG AACACTTTC GAATCTGCTG AGCAGGGTAA AGCAGAGCTC 420  
CTAACTATTT GCGAAGCACT GACTATGGA GTACGTTTAA CATCTCTAGG GAGTTTTC 480  
AATATTTCTC CAATATTTCT ATATTTCTG GCAATGTTCT ATACGAATA TGAATCACT 540  
CACTCTATCC TAAACAGAG TCTCTCTG AGTGTACTAA TTCCCAATAT GCGCAACTA 600  
CATGTGTGTC GATCTTTGG AATATATAG TATTAATATG TTTTGAACT GAAAAAAT 660

TTTACAGCTA CTGAATTTCT TATAGAGAG GAGTGGTTAG TAAACTGCAC TGTTCCTGTC  
ATAATCGAA ATGAGAGATA TTACATCTGG AGGCGCATG GCTGCTGCTT CAGTCTCTGT  
5 TTTGAGTGC AGATTTCAT TAAATGATGC CTCGTGTTAA TACACCTGGT AGATTCTGTA  
AGAGGGGCTT TATAGCAGG CTGGCAGCG CCAGCTTATA AGTTAAGGG CATTGACATG  
10 AGCGTGTAGT AGATTAATTC AGGAATATA GAGATTGTGA AGAACTAGG ACCAGCTTAA  
CTTATATGA ATGGCAGTTC TGTATGAAA AGACATTTTC CAGTCAATTA CGTGTGTTTA  
TTTAAAGCAG ACTTACATGT AAACCGAAT CCTCTCTATA CAGTTTATTT AAGATTAAT  
15 TTTATTACCG TAAAMAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA  
GAA

20

(2) INFORMATION FOR SEQ ID NO: 120:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 120:

CAGGCCCCGG CCCCCACC AGTCTGCT TCTGCCCCG CTTGGGCGCG CCCCCAAGG 60  
35 CAGGAGAA GAGCTGTCA GGAAGCTCC GCGGAGTGC AATTACGTC CAGCTGCCG  
CAGCCAGAG TTCCAGATG GTTTGCGGG GCTTGGGCTG TTCCAGAGC TGGCTGTGG  
40 CCGTCAAGCT GCTTACACC TTGGTTAGT TCTGCTTAAT TGGAAATGCT GCGTGGGCA  
TTGGCTTGG GCTGATTTCC AGTCTCCAG TGTTCGGCT GTCTATTGCA GTGGGCACT  
TCTGTGCTT GATTGCTTTA GTGGTCTGA TTGGAGCTGT AAACATCAT CAGGTGTTC  
45 TATTGTTTTA TATGATTAAT CTGTACTTG TATTATATGT TCAATTTTCT GTATCTTGG  
CTTGTTTAG CTTGACAGC GAGCAGAG GTACGTTCT GAGGTGTGT TGGAAATTA 480  
50 CCGCAATGC TCGAATGAC ATCCAGAGA ATCTAAGTC CTGTGGGTTC GAGATGTTA  
ACCAATAGA CAGCTGTCT GCTAGCTGT TTAAGAGTGA CCAGCTGTGC TCGCATGTG  
CTCCATCAT AGGAGATAT GCTGAGAGG TTGTGAGATT TGTGTGTGG ATTTGGCTGT  
55 TCTTCAGTT TACAGATAT CTGGGTGTT GCGTGAAGCTA CAGATACAG ACCAGAAAG  
ACCCGCGCG GAATCCATGT GATTCCTTT GATGAGAAA CAGAGAGAT TTCTTTCTGT  
ATTATGATCT TGTTCACCTT CTGTAATTT CTGTAAAGT CCAATTGCA GTTTAAGGAA 840

60

GGAAGACTA TCTGGAAAG TACCTTATG ATAGTCGAA TATATATTT TACTCTATGT 900  
TCTCTACAT GTTTTTTCT TTCCGTTCCT GAAAAATAT TGAAGCTTGT GATCTCTGAA 960  
5 GCTCGGTGC ACTCGGAA TTAAGTATAT CATTGTGCG CACTGTGCAC TGTGGCTTTT  
CTTACGATT TTACCTCAG AAAAGCTTGT TATGTACCA CTGTGTGCT TATATGATGA 1080  
ATCTGAAGT ACATCTCAT GTATATATA TATGTAGAC TGTGCTGTGT AGATATGTC 1140  
TACTGAAA AGATGAAA TTATTAATA TCGAAGATA TGAATCTG TATATTTAG 1200  
GGAATCCAA ATTCCCAAT TTTTGTGTC TTTTGAAGA AGATGTGTC TGTAAAGAG 1260  
15 TGTATGATA AAATGATA TTAAGTATA GTCTTTTATG ATTAACCAA TGTATTTCTAG 1320  
AAATGATAT GCTTAGGAA ATGTGCTTT AATTTTGAC TTTTACAGT AGTGCAGAG 1380  
GAGAAATGT TTCAAGAA GTCTAAATGT ATATATCAT TTACCTCAG CTTCCATCAG 1440  
20 AATGAAACA GTTTGAGTA ATCAGAGAT ATATCATAT GATCTGTATA TGTTTTATA  
ATATTTGAA GTCTAAGAA CTGCATTTT AAACAGATA GTATTAATGC GTTGGCCAC 1560  
25 GTAGCAAAA GATATTGAT TACTTAAAA ATGTTAATAT ACCGTTTCA TGAAGTTCT  
CAGTATGTA ACAGAACTT GTAAAGCTA AGCAATGTC AATATGATCT CCAATATTT 1680  
30 GAATGAAA TGTATATG TGGCTGTATA TATCTGTATA AAAAAATTA GAGAGAAAC 1740  
CTTCTTTGT GTATGCATGT TTGAATTA AGAAGTAAT GG 1782

(2) INFORMATION FOR SEQ ID NO: 121:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 610 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 121:

GTTCGCTGCA GATTGTGCT GCTTCTGAG CCGTCTGTCC TCGCCACGA TCTTCAAG 60  
TATATTAAG AACATATGTA TCCCATGAA GCGCTACTAC ACCAAGTTT ACCAGAGAT 120  
50 TTGATAGGA ATGGGGCTGA TGGGCTTCAT GGTTTATAAA ATCGGGGCTG CTGATTAAG  
AAGTAAGCT TTGAAGCTT CAGCGCTGC TCGTGTGAT CAGACCGA TTTACTTGA 240  
GTACATGTA AGAAGAGCT CAGTCTGCT GTAAATTTCA GCAAGCTGT TTAGATGGG 300  
55 AGCGTGAAC GTCACTGTAC ACTGTATATA GTACCGTTTA CTTCATGCA TGAATTAATG 360  
GATCTGAG ATGCACTGCT ACCTGATCT CATTTCAGT TGTTCGCTCT CAGCGCTCG 420  
60 GCGTGTGAG CATACTCTGA GTATATAT TGTATGAG CCGATGCAAT CAGATCTCA 480

CTGAGCCACC CACTATCTGT AATTATTTAC CTGAGTTGTA CAGGACTTGG TGATCAGGAT 540  
CGAGGCACTC ACTTGATATC TACTCTCTAA TAAAGCTTTA TTAACTTGA AAAAAAAAAA 600  
AAAAAAAA 610

5

10

15

20

25

30

35

40

45

50

55

60

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 526 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 122:  
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 122:  
GTTAGGCGCTG CAGGTACCGG TCGGATATTC CGGGTCGCC ACCGTCGAG CGAGCGCTCC 60  
ACCGAGCGGT CGGCGACGC GTGCGAGCGG ACCGGAAGTC GTACGAGATG TACAGGAGGT 120  
GCAGCTCGCC ATCTGCGCA ACATGCTGGG CGTATCGCTC TTCTTGGCTG TCGTCTCTTA 180  
TGACTAGCTG CGCTGACCA ATCCGAGAA GCAGGAGATGA AAGTGGCGGT TTCTCGCGCC 240  
CAGGGTTCCA GCACATAGTC TGAGGCGAGA TGCGAGGTAT GAGGCGGCTT CAGACTTCAC 300  
TTATCGCTTT CTAGCATCTA CACATCTACA AGCACTTACA CCGTGATTTT TCCAAACMAC 360  
TTTTATTTCC TGAGAGTCTT CATTATCTCT ATGGAACGAG AAGCTGCCAC TGATAGGCG 420  
CGAGTATGAG GCTTGTCTTT TCTACTCGCT CCGCCGATTA TAAATATATA GACTTTTAAA 480  
AAAAAAAAA AAAAAAAAAA NCGGCGGCGC GGTACCGCATC CCGCTA 526

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2081 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 123:  
(1) SEQUENCE DESCRIPTION: SEQ ID NO: 123:  
TGTACCGGTC CGGAATTCG CGGATTCGAC CAGCTCGTCS GGGGACATG CGGCTACGG 60  
AGCGCGCGGT CCGTGGCGTC CCGACAGAGG GCGCGGCGGG CGCGGGCGCG CCGTGGGGA 120  
CAGTCCCGGT GCTCTCTGT TTCTCATGCT TCGCGCGAC CCGTTCGCTG CCGACACGGG 180  
CGGCTTACA GGTGCTGATC CAGAGTTTCC TGAGCTCTTA CCGGCTCTTA CCGGACGAG ATCGACATCC 240  
ACCGGAATTT CCGTGTGAG CTGTTGCGCG AGGAGTGGGG CCAATGAGTG GACTTTGCCA 300

AGCGCTTCC GGTAGCGAG CGCTGACAGG TCGCGCTCGT CGCGTGCAG ATCCAGCTCA 360  
CTACCGTGGG AATCTTTACA CTTTCAGCA CTGTGTTTTT CTGCTGTGAT ATCCAGGAAA 420  
GGTTCAGAAC ACCATCAGG TATTTGGGG ATATATATAG CCGGCGAG AGATTTGTTC 480  
AGGGGGCGG GATTTTAGGA ATCTCTTTA TTGTACAGCA ACATTAACCT AAGGTCTCTG 540  
CGAGCAGGT TCAGGAATTT GATTTACAG GTGTAAACT GTTACTTCCA AGACCAAGT 600  
TTTCATGCT ATTACAGAA GTAGAGCGG GATTAGCAG GATTTCCGGA GTACGAGTG 660  
TTGTATATT TGAGTGAAG ACTCATGTGT GCATCCAGCA AACTCGCTTC GAGCTATTTG 720  
GCCAGGAGT CGAGTTTAC ATGTTTCTG ATGCCACTC ATCAGAGAGC ATCATGAGCA 780  
GGATGTTTCC CCGTACGCGT CTGCTCTGAT CGGGATCAT AATGACCACT TCGAGAGCTG 840  
TTCTGCTTCA GCTGTAGCT GATAGGAGC ATCCAAATTT CAGGGAATTT CAGAACTTAA 900  
TTAGCGGAG TCGTCCAGAG TCGGCTCTGC TTTCCAAGT ATAGGACAT TCGAGAGCTG 960  
GTATGCTACT CACTGTGGA GAGAGTCAG GTAGAGACT GTAGAGCCAC AGAGCTCTTT 1020  
CTTATCTCTA CTAGATTTAA AATGTTAGT GAAAGCGGC TCGTTTTCG CGGCTCTTAG 1080  
TGAACTTTAA CGAGCTGAC GATTTAGTA CGAGATTTA GTTAAAGG TCAAGGCTTT 1140  
CGGCTCTCC TTACTCTCT TTATTGTTAA TGTGCTTTTA TTATTAATA AATATTACA 1200  
TGAGAGTCC TGTTTTCTCT CTACTGTGTA CTCTGATCT ATCTTTCGA AGTGCAGACT 1260  
CTTGTAGAT TTCTTTAAT TGTTCACCTT AAGAAATG AGTACGAC AATGATTTGG 1320  
CTTTTATAT ACTGTAGAT GTTATATGT TATGTGAT GTAGTCTTT TACTTTTACG 1380  
ATTGATGGA ATAGATTTAT TGATATGAA TTACCCACA GAGCTCTGAA TCAATGTTAC 1440  
CACTCGCTC AGATGTTGT CCACTTAGTG AGTTCATG AGTATTCCT ACCAATGAT 1500  
GTGAAATAT TACTATCTTT TCTTGACTAT ACTGATTTCT TATTTTGGTC ACTATTACTA 1560  
AATCTCTGTT AATATCTCT CTTTACTG AAGAGGATG GATAGAGAG GTTTGCAATG 1620  
CGATATATT GGTGAGGCG TGTTTTACA TCTTTGAGT ATGCTTGTCT GATATCTTTT 1680  
ACCAACATCT TATATATATA TTCTAGTCT CAGAGATTT AGCAAAAGA TAAAGCTTGG 1740  
GTGGAATAT ATTTAAATAT GTTATGTC TGTCTATAT TTGCTCAC TACTCTCCA 1800  
ATATGTAAT GCAAAAGTC TGATATAGA TTGTGATGA TTAATTTGT GGTCAATGTT 1860  
TCTCTGAT AATTTATTT TCAATTAATA CTATTAGAG GATTTTGAAA TGTTTTCCA 1920  
ATATGTGAAA TGTGAGACTG CTGTCTTTA TATTAAGTA ATTAAGAAA ATGTATTTG 1980  
ATGGAATTA TTTTGCCCTC CACAGATGG CTCTATGAT ATCTCTCAG GATTTCTAAT 2040  
ATTTATTTA GGTATTAATA TCTTGACAT TATATCTTT C 2081

346

347

5 (2) INFORMATION FOR SEQ ID NO: 124:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1717 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 124:

15 CCCCAGCGGGA GCTGGACCTCG CGGTGGGCTA GGGGCGGGCG CGGCGCGGAG 60  
TGTGGATCCT TCATGATGAG AGATTGGGG ACATTTCTCT CTCCTGTGTG TAGTGTGATAG 120  
TTTGTGTGTC AGAGATGCG TCACAGTGTG AAACCTTTTC TCCAGACCT TCCGAGAGA 180  
ATCAAGAGACT CACTCTGGG TATTGTACC ATCTCAAGC TAGATGCTCG AATCCAGGA 240  
AAGAGAGAGG AGCAGGCTCG AGAAGGGCA AGTAGTGCT TGGCAGAGAG AAGAGCCGAG 300  
AGTATAGAGC GGAGCGAGGA GATGAGCCA GGTATGTGTA GTAGATTTT CAGATGTGT 360  
GCTTGGATG GTGAGTGTAT CTGCTTCAT CTCCTCTGT TTTATCGAGT ATTATTCCT 420  
GTGCTTCACT CGATTACGC CCGATTATC GGTGACCAAT CACTACATG AGATTTTGG 480  
TGTGTGCTCG AATCTTCCT CAGCTCAAT TTCAGTGTCT TTTGGTCTC CCGCTGTGT 540  
GTGCTTAGCA AATGTGTGAA TGGATTTGG TTTGAGGATA TAGCTGACT GGCATTTGAG 600  
GTATCAGGGA GGAGGCTCA CCAATTCCT AGTGTAGCA AATATATTC TGCATGCTC 660  
TTCAAGCTTT TGTGTGAGC TCTTTTCTC ATTCAGGGA TGTTTGTGAG TCTCTTTCC 720  
ATCCATCTTG TCGGTGAGCT GGTTAGTCTC CTGCATATGT CCGTCTCTTA CTCAGCTTAC 780  
TGTATTGAT ATCTTGTGTT CATTAGAGA ATTGAAATGC ACAGGGGTT GTCTAACATA 840  
GAAGGGAATT GGCCTTACTA CTTTGGGTTT GGTGTGCGCT TGGCTTTTCT CACAGCAATG 900  
CAATCTCAT ATATATTCAG TGGGTGCTT TTTCTATTC TCTTTCTTTT ATTCATATTC 960  
AGCGCCAAATG AAGCAAGAC CCGTCGAAA GCTATATCT TCCAGTTTGG CCGTCTTCTC 1020  
TTTGTGTGCT TCTTAGGAA CAGACTCTTC CACAGAGAG TCTACTTCCA GTGCGCCCTG 1080  
AGCAGCTCTA CTCCTGAGA GAATTTCCCT TCACCGCATC CCGTCGCTCG CAACTGAG 1140  
GCTATCTCAG GTACTGATGT TGGCTGCGAT CCAAGGGAAG TGGGCGGGAT TGGAGAGGC 1200  
TGTGCGAGCT CTTTTCCTG TTCAGCTGCC GCGTGGCAGG GAAGGCGGA CCGGCTCTCG 1260  
CAGGGGCGCT CTCGTATTC CTTCTCTCT GAGGATTTGA AATTTTATC TCTGTGTGAC 1320  
GTAGGCGAGA ATGTGCTCTG ACAGCTGT GTGGATTTT AACATCACG TGAATCTGAA 1380

AGGACGACAG GTTTTCTGC AGCTATTTTC TAGCAATTCG CAATCCCTGT GCTGTGACTG 1440  
ATTGGAGAC ATTGTATTTC TCCGTGTGCC ATTTACCTTT CCACTTTTCC ATCTGTCTTT 1500  
CTACACCTCT TGGATGAATG GATTTTGTAA TTCTAGCTGT TGTATTTTGT GAATTTGTAA 1560  
ATTATTTGT TTTTCTGTCA AACATATCA TTGGATATG GAGGTAAAGG AATGTGCCAG 1620  
TTGCTCTCTG TCATCTCTTT TATAGCCATT ACTGTCTTGT TTCTGTATAC TCAGGTATAG 1680  
TTTGTGCTCT TCTTGTCTCA CTCGAAAAA AAAAAAA 1717

(2) INFORMATION FOR SEQ ID NO: 125:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 804 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 125:

CCAGCGTCC GGTCACTATG TATGTGGGG GCGAGACCC TCCCGCAAT TCTGTGAAGT 60  
TCTTATGCTC GACTAGGGA GTAGCCCGAG GACTCTTAT GCGCGGCTTC AGGTCACTCC 120  
CGGCTGAGCG GAGCTGCTCT CCGATTTT GGTGTCTTGG TGGCGGGGAG GCTGTGTCAA 180  
ACAGCTGAC AGCAATGGC AGAGATAAA TTTGTTTTTG ACTTACTCTA TTTATGAAT 240  
ATCAACATG TTTGTGTTTT TATGTGGGA ACATCCCAT TTCTGTAGGG AATGGAGGA 300  
TCTGTCTACT TTTCTTATCC TGAATTAAT GGAATGGAG TATGGCACT CTTAGGATTT 360  
GTCAAGATG GGAAGCCAG TCCATCTTC AATTTTCTAG GTCTTAAATC TGGAGAGGA 420  
AGCCAAATC CTTTGTGAGC CATGATATAT GTCCGAATC CATCTGTTC TCAGATTGGA 480  
ATTCTAGTGG AATTATAGA CAGTATGCT CAGGAGACT CTGTAGGTAA TGTCTGTGTA 540  
TCTCAGTTG ACTGATCAC TCAGTTTACA CAAGAATGT TGGCAATTT CTACATTTT 600  
GCTTCATCAT TTGCTGTCTC TCAGGCCAG ATGACAGCA GCGCATCTGA AATGTCTATT 660  
CGCGCAATG TGTGTCTGAA ATGATATGAA AACTTTCAA GAGGACTAGC ACAGAGCCT 720  
ATTTTGTGGA AACATTAAT TGAATTAAT AATTTTAAAT GATTTATGTA AAAAAAANA 780  
AAAAAAAAA AAAAAAANA AAAA 804

(2) INFORMATION FOR SEQ ID NO: 126:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 431 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 126:

5 GGCACAGCCC AGGCGCTTGA AGCCAGCTGG CCTGTGGAG AGGCTGTCTGT CCGACCTTGG 60  
GGAGGCTCTG GCGTGGGCGT GCGCTGTGAT GCGGTACCTT CCGAGCACTT 120  
GCTGGGGTGA ACTTTATTTT AGCCCTTCCC TTGTGTCTCT TATGGAGAA CAGAGAGGG 180  
GTGGCGAGGT CAGTGTATCT ACAGATGGAG TGAATTCGAG CACAGCGGCT TCTGGGAGA 240  
GGGATGGAG GCAATTCCTTT CAGGGAATG GTCCATNAT TTACCCAGAA GGCATTCGAT 300  
TAAATTAAAT CCGAGCTTT TTGTGGCGAG CTCTGTGTTA TTAGGGGCCC TTGGCGAGA 360  
CTTCAGGAG GGGCGAAMAH GACCTTTAAG TTTTAAAGTT TAAACACAGG AACCCGAAA 420  
GGTTATTTT G 431

(2) INFORMATION FOR SEQ ID NO: 127:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3752 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 127:

35 NCGACAGGG AAGTCACTT GAACTAGAA CTAGAGATAT CCAATGACC AGAAMAMTT 60  
AACTTCAGC TTCTTAGCA TAAGGATTT CAGAGACTC TTGGTGGCA GAGCGCTGTG 120  
TATGATACCA GATTAAGAAC TGGCAGACA CTGAAGAAA AGACTTTTCT TCCGAGAT 180  
ASTCAGAAC TTGACANTTT CTAAGGAAA GTACAGACA AATGGGATAC TGTTTTGGC 240  
AAGTCTGTGG AGCGGAGCA CAGTTTGGAG GAAGCCCTGC TCTTTTGGG TCAATTCATG 300  
GATGCTTTGC AGGCTTGTGT TCACTGGTTA TACAGCTGG ACCCAGACT GCTTGAGAC 360  
CAGCCCTTGC AGCGGGACC TTGACTCTGT CATGACTTC ATGGATGCAC AAGAGGTTTT 420  
CCAGAGGAA CTGGGAGAG GCAACAGAA CCGTTCAAGT CCTGAGCGG TACAGGCCGAG 480  
AGCTGATGTA GATTAATGCA GATACACCA CTTGGGTAAA AGCAGAGCTC CAGGAAGTGA 540  
GCATCTGCTG GGAAGCTGTC TGTAACTCT CTGTTTCCA ACAAAGCCGG CTTGAGCAG 600  
CCTTAAMACA AGCGGAATG TTTCAGACA CACTCCACT GCTGTTGGAG TGGCTTTCTG 660  
AAGCGAGCA AACCTTGC TTTCGGGAG CACTTCTGAG ATGACAGAA GCGCCCTGAG 720  
TCTCTCATTG ACACCCATPA GCAATCTG AGCAATGAG AAGAAAGCG AATGAGCTT 780

5 AACTCAGAG TAGCCATGG AGAATGATC CTGGCTGTCT CCGACCCGCA TTGCATACA 840  
ACCATTAAC ACTGATCAC CATCATCCGA GCTGCTTGG AGAGGTCTCT GACATGGGCT 900  
AAGCAGACC AGCAGGCTCT TGAAGCGCC TTGTCAAGAC TGTGTGCTAA TGTCTAGCTC 960  
CTGGAGAAC TTCTGGCATG GATTCAGTGG GCTGAGACCA CCGTATCTCA GCGGATCAG 1020  
GAGCTATACC CCGAGACAT TACCCGATTT AAGAGCCCTTA TCGCTGAGCA TCAGACATTT 1080  
ATGGAGGAA TGACTTCCAA ACAGCTGAC GTGAGCCGG TCACCAAGAC ATACAAAGG 1140  
AAGAACTAG AGCTACTCA CCGCCCTTTC ATAGAGAAAT CCGCAGCGG AGGCGGAAA 1200  
TCCCTAATGC AGCCAGCCC TCTTCCCATG CCAATCTCTT CAGATCTGTA AGCAAAAAC 1260  
CCAGGATCA ACCAGCTTTC TCCCGCTGG CAGCAGATGT GCTGTTNAG ACTGAGAGGG 1320  
CAAGGAAC TGAATGATGC CTTCATGCG CTGGAGAGT TGAAGAATTT TCCCACTTT 1380  
GACTTTCATG TCTGAGGAA AAGTATATG CTTTGATGTA ATCAAAAA GTCTCGAGTG 1440  
ATGATTTCT TCCGGGCAAT TGAATAGGAC CAGGATGGGA AGATACAGG TCAGAGATTT 1500  
ATCGATGCA TTTTATGATC CAGTTTCCC ACCCAAGAT TAGAGATGAC TGTGTGTGCT 1560  
GACTTTTGG ACCGAGATGG GATGTTTAC ATTGATTTT ATGATTTTGT GCTGCTCTTT 1620  
CATTCACCA AGATGCTTA TGGACACCA ACCGATCGAG ATAAATGCA AGATGAGTTT 1680  
ACAGAGAG TGGCTCAATG CAATATGTA AAGGTGTTT AGGTGAGCA GATCGAGAG 1740  
ATTAATACC GGTTCCTTCT CCGCATCAG TTTTGGGATTT CTAGGATTT GCGGCTGTC 1800  
GTATATCTGC GCAACCTGTA TGTTCGCTT TGTGTAGGA TGGATGGCT TGGATGATTT 1860  
TTTATGAAA AATGATCCT CCGAGCAGG AGGTAGAGCT ACATTTGAC TTAGAGAGAA 1920  
ATTCATCTTA CAGAGGGAG CATCCAGGG AATGACCCC TTCCGCTCAC GGGGTGAGG 1980  
GTCCAAACA TCTTCCGCG CAGCTTCCC TACTGCTTGC AGCTCCAGT CTATGTCAGG 2040  
TAACAGAGC TGTACATCA TCCATCTTC TCCAGCCACC CCAAGCATG GAACCAAGT 2100  
TATCCATCA TCAAGTACA AATTGAAAG ACCACACCA ACTTTTCAAT CTATGTCAGC 2160  
ATCCCTTCT GGTGATACA GCAATGATTT CTTTCCCGGC CTCACAGT GCCAAACTA 2220  
ATCGGCGAG CCTTAAGAG TGTGCGAGT GCGCTTGGAG TCGGCTTGG AGTTCAGCG 2280  
GGATGAGC CAGAGCGCG CAGAGATG ACCTTCTGTA CTTTGAAGCT TTAGAGAGC 2340  
ATTGCTTTT CCGACTTTC AAGAGAGC GTTGCAGGG CCGAAGGCA CTCAGAGAGA 2400  
GGCTTAACA AACCTTCCA ATCTCCACC ATGCTTAAGA AGACACAC TCCCTCCGC 2460  
AGGATCGAG GTCCGAGCG ATACACTGT CTAGCAGCC CAGGCACT ATGCACTTTG 2520  
AATCTGCTC CATACATGG GTGTATATT ATTCTGAAG GAGAGATTA TATGTTAAA 2580

5 AGCTGAAAG AATATTTGTTT TTAAGAGCT GCTTATTTT GTAAAGTACT 2640  
ATTTTCATCT GATATTTAT GTAGATAAA TTTGCTCTCT GTAAAGCTTG TAATGAGATG 2700  
GCTCCAGAAA TGAATATTTT GAGAAAGAA AGTGAAGAG TCAAGATACA AATGTATAT 2760  
AAAAAAAAA AGGCTATTA ATAGGTTTTC TGCGCGGTGC AGGGTGTGAA AGCTGCTTTA 2820  
TCTTTTAGA TTATCTTAA ATGCATCTTC TTATTAAGT TGACTGTGTA TCTCAGCAG 2880  
ATTAATATA TTAATAAAT AGATATCTTG CAGTGTGTA GGAATCTTTT TTTTGTAAAT 2940  
CAGCGAGCC TCATATAGA AGAATGAGG GAGGGCTTT TGCATTTGTT TAATGTTTTC 3000  
TGATTTTTCG CTAAAGAGAG GGAATCTCAT CTAAATACA TTTGCAATG ATACAGCAA 3060  
AGAGTTTCAT TGCATATCTG TCTTTGGATA TTGTTTCACT ACTGGGTGTT TAAAGGACA 3120  
ATAGCTGCTA GAATTCAGGG GTAAATGTAA GTGTTCAGAA AACGTTCAGAA CATTGCGGTT 3180  
TTTAAATGTA TTTGTTGCTC CCAATCCAG CTAGACACCA GTAACTCTTG TTTTCACCG 3240  
GACCGAGCC CTTCGAGAG GATAGGCTCG TTGTGATAT TGTGAATTC AGATTTGTTT 3300  
TATCCACTTT TTTGCTATTT TATTTAATG GTGATGAC TTCCACAAA CTGAGGATG 3360  
AATTCAGCA CCTGTCTGCT AAATGTGGA CTAAAGACA AGAGTGTCTC GTCTTTTAT 3420  
GGAGTTACG AGCACTAGT TTAACAGTC CTGTTGCTT TGTCTCTTTT TTTGCTGTA 3480  
TAAATGTCAG TCAGCATGTT ACCTGAAA GGGGCTGCTT GGGGCTCTTG TTTTTFACCT 3540  
GCTGTTCTTC AGCTCGAGCC ATGTGCTGCT GTGATATCT CAATGCTTTT TAATTCAGCC 3600  
AGAACTGAA CCTTCACCA TGAACTGTTT AGAAACAGA CAGACTTTTG TATTAAATTT 3660  
GCTTCAGTA ACAAAAAA AAAAAAAA AAAAAAAA AAAGTCAGG GGGGCGCGGT 3720  
ACCAATTCG CCTATATGA TCTAAACA TC 3752

(2) INFORMATION FOR SEQ ID NO: 128:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1144 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 128:

55 TGAACCTCTG CCGCGGCG TCAGTGTGAG AGGCTTTCTG TTTTGTGCA GTGCGTCTC 60  
GTAAACCA GCGGCTGTG GTCAACCAT CATTACGA GCTTCATTTG GTTCAGCAAC 120  
CTTAGACG CTTGCTCTT ACCACTCCAG CAACAGCT GCAACCATC GGCCTCATG 180

5 GCACAGAAA CTGCTCTGCT CTCAAGCTC CTTGCACTC CAGTCCAGCT GACGACTTGG 240  
GACAGACTA CAAACCTGAG CGATATGGA GCTCCAGAT TGAAGAGAA AGCTCGAGCC 300  
CTTGTCTTAA TTGCACTTT CTCAAGAGA ATTAATTTAA GCAAAAACCA AACAAACATA 360  
GTGCGCTCTC GTTAGATCA TATGTGTCCA GTTCTGTAGA CATCTTTTAA AGGCTCTTAC 420  
TGCAGCTCC CTGCGACCC TCCCTCTCTT TCCAAAGAG ACCGAGCAG GGCAGGCTCA 480  
GACGACTGC TTTCTTCA TAATTCTGCA TCTTCTGTC AATTAATATA AATTAATAT TCTCTTTG 540  
CTTTTAGGA AAATGTGGA CTCAAGACA AACCTTAATA AGACTGTATC ATCTAGAAC 600  
CTTCCGCTT ACTGCTTTT CAGCACTCT CTGCGCAGT GTTTATTTAT CTGTATGAC 660  
ACAGATTTGA CATTACAGCT AAGGAATTA TTTGAGTTGA TTTCAATTC CTGCAATG 720  
ACATTTTCT TAATTAACA AGTTGCTGTT TTAATTAAT CTCAATATA TCGGCCAGA 780  
TCTAATTTTA AAATGTATG AGCACTTTCT GCTAAATAA GATTAATTTT TTAAGTATG 840  
GCTGTCTGCT TTTAAAGCA GATTAAGAA ATTAATGTA ACTTAAGAC ATTAATGTA 900  
TGTAAACCA TTAAGTATG ACATATGCA TTTTCTGTC TTTTCTTACT TGAAGTAT 960  
ACATTTTGA TACTGACT TATTTTAAG ATGACTGAA ATGCAATAG CAAATCTCTG 1020  
AGATACAGA TTAATGCTCT ATTTCTTAA AATACACTT TGTGTGTATC TTTGAAATTA 1080  
ATGATGCTTT TTTCAAAA AAAAAAAA AAAAAAAA TCGAGGCGGG GCGCGGTACC 1140  
CAAT 1144

(2) INFORMATION FOR SEQ ID NO: 129:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1810 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 129:

50 GATTCAGAG GAGCACTG AGGTTTCTC TGAAGAGCC GGCATTTTG GCACGAGCC 60  
ACGGTTTCT GCGCAAGTC ATCATCAGG CCAAGGATAT CATTCGAGAC AGGCGCCAG 120  
GTGAGATTT CTGCGCAAA ACTTCGAGT CAAGACAG ATGCGCCAGG GTGCTCCGG 180  
CTCTACGCG CTCTGCGAG CGCGGTGGA TGGGAGCTC TGAATGCCA AATGCGCCAC 240  
GCTGCGCGCC GCGCTCTGGA GCTGGAATTT GGAAGAGAC AGCAAGCAC GCTGCGCTTC 300  
TCAAGGATG GCGCAAGCT TCGCAAGCG CCGTTTCGG GACTTCGCA GCTGCTCTCC 360  
TGCCTCTTC GCGAGAGC CTGCGACCC TGCCTGTGAT GACGAGCTGG CTGATTCGCC 420



CTGGGCGGCG CAGTCTCTCA CAGCCCTCCG AGAAGCTGGA GGGGTCTCTGG AGACCACTAG 480  
AGCTGATGGG AGCAGCTGGT GCGTGGCGTT GCGCTCTCCG GTTCCGAGAA CCGAAGGGAA 540  
GTTATGAGAG GCGAATAGGG GCGACCGCGG TCCCTCTGGGA TGGCTCCGCT GCACTTTTGA 600  
AACCCCGGTT TCCCTTCAGG TCCCAATCC AGGTGACAC AGCTGTCTCC TCCCTCTCAT 660  
CTTAGCTTCC AGGTTCACCC TAAACCTCTA CTAACTCTCT TGGTGGACTT GGAAGAAGCT 720  
TGGCTCTCTG GGAAGAGAG AGAGCGGCGC TCCATCAGCG CTGTTACGAG AGGATCCCGG 780  
AGAAGCAAC CAGCTCTGGA CATCAGCGCC CCTGGAGCTG GGGCCAGCAG CCGTGGGZAC 840  
GAGATTGCTC CTGACTTTAT TTATATGCGA TGAATCTCT GATTATTTT GGGATTTTTT 900  
GTTGTTGCTG TTGTCAGT TTGTTTTTTC TAAAGTTCTG TGAATATATA TTTGACATTT 960  
TACATTTCAA AGAAGGAT GTTGTCTAAC AGGGAGCGAA CAGAGGCTAG TATTGACAC 1020  
TGTCTCTCTC TCTACTTAAA AAAAAAGAGC ACAAAAGAAA AACTAAATTA TTGAAATTT 1080  
AAAAATGTC ATTGTTCTCT GTTGTATAT ATTAGGGTTG TTAGGCTCTG TTTTGAGGTA 1140  
TGGACTGTA TTGCTTTCCC CAGCCCTCAT TCTTCAGCGG TTGGCCGGTG TTAGAGCTCG 1200  
CTCTCTTCTA GTGACTGGCT AGAAGGCGCT GAGAGGTGGC CAGCCAGGCT TGGAGCTGGA 1260  
GGGAGTGGAG CCCCACCTGA GTTCCCGGTT CACAGCGGTT AGAGGGTTCAC TGGGAAGCAC 1320  
CGGCGGCTGG CTCTCTGTAT TTATTTTCTT GATGTAACT TCTCAGAGCA GGGCATTTGG 1380  
GACATACCA GCGAGAGAC AGAAGGACAC CTTGCTCTCT GTGGAGGAGG GACCCAGACA 1440  
AGCCCGCTCG GCAATTTGTC CCCCAGCTTT GGTATTCCT TCAAGGAAAG GTACAGCTG 1500  
GGGAGGAGCG GGGGGAGCGC CTGTACGCC TGGCAGGTGG TGAATTCAGG TGGGGGCTCC 1560  
CTGCTACGCC CAGGCTTGG AGCTTGAGC CTTCCCGGCA TCTGGCATCC GAGCCCTCGG 1620  
CCCTCCAGGG TGGCTTTCC TCTCTTGGG CAGCATACAC GAGGGCAGGC AGTGGGCTTG 1680  
TCACTGTATC TTGCATCAGA GACAAAGGAG GACCCGCTTT AGCCCTCTCT GGGGAATGG 1740  
GGGATGGGCC AGGGGAGCG CATTGTGCAC TGGTTTAACT TAAATATGAC AGATTCTCT 1800  
GTTAAATTC TTGATGAT TTTTATAT

(2) INFORMATION FOR SEQ ID NO: 130:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1864 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 130:

GGCGGCGGAG ATGGGACCC CAGCTCCGC CCGAGACACA GGGGTCTCTGG TGGAGACTTT 60  
5 TGTAGTTTAT AAGCTGAGCG AGAAGGTTTA TGTCTGTGGA GCTGGCGCCG GGGAGGGCCC 120  
AGCACTGAC CCGCTGCACC AAGCATGGG GCGAGCTGGA GATGAGTTGG AGACCGCTTT 180  
CGGGGGCAGC TTCTCTGATC TGGGCGCTCA GCTGATGTC ACCCAGGCT CAGCGCCACA 240  
ACCTTTCACC CAGGTCTCCG ATGAACTTTT TCAAGGGGGC CCGACTCGG GCGGCTTTGT 300  
AGCTTCTTTT GTCTTTGGGG CTGCATCTGG TCTGAGAGT GTCAACAGG AGATGGAACC 360  
15 ACTGGTGGGA CAGTGGCAGG AGTGGATGTT GGCCTACTG GAGAGCGGCG TGGCTGACTG 420  
GATTCACAGC AGTGGGGGCT GATTATCCCA GATCACTGAA GCTGAGATGG CTGATGAAGT 480  
AATTTGCACT GAATTTTAAA GCGACTGTGA CTCTCTGCA AGTTCCCGAG ATCTTGAGCA 540  
20 CCTGGAGCT ATCAAGCTC GAGTCAAGGA GATGAGGAA GAGCTGAGA AGCTTAAGCA 600  
GCTACAGAAC GAGTGAAGGA ACCGATGAA TATGCTTCA CTTCCAGGCA ATGCTGGCCC 660  
25 GGTGATCTG TCAATTCAGG AGAAGATGGA GGTGATGCG GATTCCATCT ATGTTGGCAA 720  
TGTGACTTAT GTTCCAGCAG CAGAGAGCT GAGAGCTCAC TTTCTATGCT GTGGTTCACT 780  
840 CAAACGTTT ACATACTCT GTCAAAAT TTAGTGGCAT CCGAAGGCT TTCGTTATAT 900  
AGAGTTCTCA GACAAAGCT CAGTGAAGAC TTCTTTGGCC TTGATGAGT CCTATTTAG 960  
AGGAAGCAA ATCAAGGTGA TCCCAAGCG AACCAAGCA CAGGCACTA GCAACAGCA 960  
35 CCGGGTTTT CAGCGAGCCC GCTACGGCG CCGGACCAAC AACTACAGCA GCTTCCCGCTC 1020  
TCAATCTCAC AGTGGTTTTA ACAGAGGCC CCGGGTCCG GTCTACAGGG GCGGGGCTAG 1080  
AGCGACATCA TGGTATTTCC CTACTTAAA AAGTGTCTA TTAGAGGAG AGAGAGGAAA 1140  
40 AAGAGGAAA AGAGGAAA AAAAAAGAT TAAAAAATA AAAAAAATA ACAGAGGAT 1200  
ACCTTGATGG AAAAAAATA TTTTAAAAA AAAAAAATA CTGTGGAGG GGGGAGATC 1260  
45 CCACTACTTA CTGCTGAGGA GGGAGCTCT TTGGGAGTA GGGGAAGGCC CAGGGGATGG 1320  
GGCAGGGGCG TGGTATTTCA CTCTGGGAT TGGCATGGA CAGCTCTCAA CTGCGCATCT 1380  
1440 GCTTGGCAT GTTTCCCTCC CCGACCCAC CCGCTCTCTC GCGCTCCCTG CCGCTCCAGA 1460  
50 TTGCTGCTGG ATCTATTTTG TTTCTTTTGG TGGTTCTTTT TCTGTTTGA GTCTCTTTCT 1500  
TTGAGGTTT CTGTAGCGGG AGATCTCCG TTCCGCTCCC AGCGGCTCCA GTGTAAATTC 1560  
55 CCGTTCCGCC TGGGGAAATG CACTACTTGG TTTTGGGGGG TTTAGGGGTT TTTTGTGTTT 1620  
TCAATGTTT TGTTTTTTGG TTTTNTTNT TTTCTTTTCC GTTTTTTCC TTTTATTTGG 1680  
AGGGAATGGG AGGATGTCGG AAGAGGAGG TGGGAGTGG ATTTGTTTGA TTTTATTAGC 1740  
60

5 TCATTTCAG GGTGGGAT TTTTITTA TATGTCTCAT CATTAAGTT GTTTTGGAA 1800  
AAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1860  
AAA 1864

## 10 (2) INFORMATION FOR SEQ ID NO: 131:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2041 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 131:

20 GCGAGAGCG GCGGCGAGG CCGTGGACC GCGCGGCTCC GGGGATGCT GAGCAGGGG 60  
CTGCTGGCC TGTGTCTGC CPTCAACCC AGGAGATGA AGCTGCTGT GGGCAGTCC 120  
TTCTGGCT AGTTCCTTC TTTTGGGCG ACTTCTGTTA ATATGAGTC TATCAGGAA 180  
AATGCGAAC TAAATATGA AAGCAGATT GAGGAGATG TGGAAACCT AGAGAGAAA 240  
ATCAGAGAT TAGAAAAAG CTTTACCGG AATATCCAC CAGTAAAGTT TTTATCGAA 300  
AAGATCGGA AAGAAATTT GATACAGGA GCGCGAGGCT TGTGGCTC CCATCTACT 360  
GCAAACTCA TGTGCGGCG CCAGAGGTC ACCGTGCTG ACATTTCTT CAGCGGAGG 420  
AAGAGAGCG TGGAGCACT GATCGACAT GAGAACTTG AGTTGATTAA CCAGAGCTG 480  
TGGAGCCTT CTACATCGG GTTGACGGA TTATCAATCT GGAATCTCA GCTTCCCTC 540  
CAAACTACAT GTATATCTT ATCAGACAT TAAAGACCA TACGATGCG ACNTTAAACA 600  
TGTGGGGCT GCGAAGCA GTGCGTCCC GTCTGCTCT GCGCTCCCA TCGAGGTGT 660  
ATCGAGATC TGAAGTCCAC CTTCAAGTG AGGATTAAGT GGGCCAGCTG AATCCATAG 720  
GACCTGGGC CTGCTAGAT GAGGCGAAC GTTGTGCGA GACCATCTC TATGCTTCA 780  
TGAGCAGGA AGCGGTGGA GTGCGATGG CCGAATCTT CAACACCTTT GGGCCAGCA 840  
TGCATATGA CATTGGGCG GTAGTCAGCA ACTTCACTCT GCGGGCTCT CAGGGGGAGC 900  
CACTACAGCT ATACGATTC GGTCTCGA CAGGGGCTT CAGTACCTG AGCGATCTAG 960  
TGATGGCTT CTTGCTCTC ATGACAGCA AGTTCAGGAG CCGGCTCAC CTGGGGAGC 1020  
CAGAGAGCA CACAACTCA GAATTTGCTC AGTTAATTAA AATCTTGTT GTTAGCGGAA 1080  
GTGAATTTCA GTTTCCTCC GAGGCCAGG ATGACCCACA GAAGAGAAA CCGAGCATCA 1140  
AAAAAGCAA GCTGATCTG GGTGGGAGC CCGTGGTCC GCTGGAGGA GGTTTTAAACA 1200  
AGCATTTCA CTACTTCTCT AAGAACTG AGTACAGGC AATTAATCAG TACATCCCA 1260

5 AACCAACC TCCAGATTA AAGAAAGAC GAGCTGCCA CAGCTGACT CCTCACTTT 1320  
AGGACAGAG ACTACCATG TACACTTAT GGAATATAT TTGCTTTT TTATTTCTC 1380  
GTTTAAAGA AACTTTTAC AGGTCTCTG AAGAACAAAC TGAATTTCA TTCTGAAGCT 1440  
TCTTTAATG AATGATAT GCTTAAAGC TCCCTCAMA AACTGCGA TTATGCTTG 1500  
CACTTTTGA ATCTCTCTT TTATGTAAA TAGGTAGAT GCACTCTCC GTATTTTCA 1560  
GTTTTTTAT CTGCTCTGA GACATATCT TGTACTCTC GTTCAAGTT TTATTTACTG 1620  
GTTTCTTGT GAGCTGAAA AGGACATTA AGCGGAGAA AATATGCCA TTTTATTTAT 1680  
AAAAATGCT ACTTAATAA TGAATCTTA TACTATCAT AAGAAAT CTTAGCATTA 1740  
TTTCTAGTG GTGCTGCCC GGAATGAT TTAGGCGGA TAAAGAAAT CTGTGTAGA 1800  
GCTTATGTT TCTCTTTAA TTCAAGTTT TTCCAGGTC TACTTTTGG TTCCAACTT 1860  
GACTTGAAA TTATCTCTT GTCTATGTC AAGGATTT GAATCTACT CTGTCTTGT 1920  
CTGCTATCT GCGCGCGCG CAGTTGCGG GGCAGAAAT TACATATTC TTGCTTACC 1980  
ATGTTAATAT ATGCTATTTT AATAAATAT TGAATCTAC CAATAAAAA AAAAAAAA 2040  
A 2041

## (2) INFORMATION FOR SEQ ID NO: 132:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2012 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 132:

60 TACCAAGCTG CAGAAATCTA CTATATCATG GCGAAGAGAG TAGATGCGA CTATGCGCT 120  
GACCGAGCT GCGAAGCGGA ATGCGCAAC CAGTCTGGA AGGACATTA TGTATCATT 180  
TTCTGAGCA AAGAGATGG GCTCTGCTT TCCAGTACA AAGAGCTGT ATTCAAGAA 240  
TACATGATG GTACATTCAG GTTCTCTGG CCGAGAGCTG GACCGAGGA AACTTGGGA 300  
ATCTTGGCT CACTTATCA AGGTGAAGT GGTGATATC TGAATATTC TACTCTCTG 360  
AATGCAAGC GCGCTACTC TGTGATCT CATGAGTGC TAGAATTTAC TACTCTCTG 420  
GCACTGCTG CTAGGCTGG TGAAGTGTG ACTTATCATG GGAATATCC AGAGAGTCT 480  
GCGCTGCGC CAAATGACT GTTGTGTGT TGTGATCTA TTATCTTGA GTGATTTCA 540  
TGAAGCAT GTATATGTC CTGTGCGGC CTGTGCTAT CTGCAAGAG GCAATCTGG

5 NAGCCGCAATG GAGGACGAG TCAATGAGT CCGGAATTTG CATTTGTTT CTTCATTTT 600  
GATCAAAATA AGCTTTGGTA TTTCAGGAA ATGTGGCA CCATGGGTC CCAAGATCCA 660  
GCCGTATATA ACCTACAGGA TGAATTTTC TTGAGAGCA ATAAATGCA TGCATCAAT 720  
GGAAACTT ATGCCACT TGGGGTCTT ACCATGACC AGGAGAAC AGTGGGCTG 780  
TACATCTGG CCAATGGCCA AGATGAGT CTACACACA TCCACTTTCA TCGAGAGAC 840  
TTCTCTATC GGAATGGCA GACTACCG GCAGATGAG TGGATCTGT CCAAGGACT 900  
TTTGAGTTG TCGATGAT GCGAGACAC CTCGGACAT GCGTATGCA CTCCATGTG 960  
15 ACTGACATG TCCATCTGG CATGAGAC CTCTTCACTG TTTTCTCTG ACAGAACAC 1020  
TTAGCCCTC TCACTCTAT CAGGAAGAG ACTGAAGAG CAGTGGCCC CAGAGACAT 1080  
GAGAGAGCA ATGTGAGAT GCTGGCAGT CAGATCCCA TAAAGATGT TCGATGCTG 1140  
GCTCTGTT TGGTGGCAT TATGTACAC CTTCCTCTG TTGTTCTGG TCTTGTGGA 1200  
GTCTTTTGT ACCACATG ACAGAGAG CTACAGGCA ATAGAGGTC CATCTCGAT 1260  
25 GACAGTTCA AGCTTCTC TTTCAGAG TACATCTG AGCTGAGAG TATCTCTAG 1320  
AAGCACAT CTATGCACT CCAAGCAGC CATGCACTAG TCACTAAC CCACTCATA 1380  
GGGCACTG TGTGAGGA CAGAGAGAG CATCAAGCT TATCTGATA TTCTTTCTT 1440  
TATTTATTT ACATGCAAT ATATGATTT CACTTTTCT TTAGTTTCT TGTCTACGT 1500  
GGGCACTG CACTAGGGA GTACTTAT ATCTACAT GCAATTTCA AGCTTACT 1560  
35 TATATTTCT TCTGACAT GAGGTATT GAATTTCTA GAATGTATC CTCTACAA 1620  
ACTAGAGCC AAGGAAAA CTCATGAT GGTTTCTAC TTCTTCAG CACTCAGGA 1680  
ATTCACTTT GACTGAGC CAGTACGT GTTAGATTA CCACTTTA ACTAAGGC 1740  
TAGCATATA GCTTATG GAAATGAG GTAGGCTGAG TATGGGAT CCAATGGA 1800  
TTTGTATCT CTTTGGAG GACTACTTT GAGGAGTGG TCAATGGTT GTTCTGCA 1860  
45 TCGCATGTA CAGCTCTG AGCTAGAG TCTCAGGA AGCCATTT CCAATGTTT 1920  
AAGCTGCG ACTGAAGA ATGTGATTT ACCTTTTAT GTTTTAGCA GGAACCTTA 1980  
50 TCCATTAAG TACTTTTAT AAGTGAAG AA 2012

(2) INFORMATION FOR SEQ ID NO: 133:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1669 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 133:

5 GAGCAATTT TTAACCACT TGTATTACAG ATGTACAGT TCAATTTAGG AATCAGAAA 60  
AGACTTTT TGTCTTTT TCTGTATGT GATCATGTT TTGTGGGTC TTCCATGCA 120  
CATTTACTG TTGCTGGTC CAGATGTTGA GGGCAATCT AGGCTGACAC ATCTAACCG 180  
AGGACAGCC TGTCTCCAT TTCTCACTC TCCCTGCCC ATATAGAC TCTCCAGGT 240  
TTAGATTACC GTTTTCGAGC ACHATTAC CAAAATGCC CCAACAGGT TTTTATCTG 300  
TTATATATA TACTTTTAC ATATACAGC CTAAATTTTA TTATTTTGT CTCGCCCAT 360  
CTGATACAA ATGTTTAAG TTGTTTAAA TCCAAATG GTAGTTTCA TGGTTAATA 420  
TTTCTAGCC TATGTAGAG TTACAGCCC ATAGCATAGA ATTAATCAG TAGCATCTGA 480  
20 GACTTTTGA GCACTAGGG CTTCTCTGG CCTACAGCC TCACTTTCCC AGCTCAGCT 540  
TCTCTCTCT TCACTCTCC ATCAGGCTG TTATGGCAC CTGTATGAG CCAAGTCTG 600  
GTCCAGGGA ACHAGAGG TTATGCTC TCCCTAGAC TATGAGTC AGTTTAATTC 660  
25 ATGATGAC ATGATTCAT TTTATTTTT ATTAGCTTT CTTAGACATA CCAACCTC 720  
ATTGTAAT CAGTAAAT ATTCAGTTT TGTGTTAGA AGCTAATA TGTGTAGCT 780  
30 GAACAAAA TGAAGTGT TTCTCTCTG TTATCTAGA GTTCTAGTT ACACATGCT 840  
GATTAATTC ATGTTCTAG GGCCTTGGC ATCTGCCAT GACTGATCC CAGGAGAAA 900  
AGCCAAAG GAAGCCAG ATCTCTTGG AGTAGATG GGAAGAGCC CATTCAGGA 960  
35 TATGAGTCC TGTGAATTC AGTTGTGTT GTGCTCTCT GTTAGAGTC ATGTTGACT 1020  
GCTTTAGA GCTCTCCAT CCACTTTTA CATGATGAG GCAACATGT CTTGTAGAT 1080  
40 TAACTTTGG ACNATGAG TTAGCTTGA GAATATGGA GGCCTCTCT GAGCCAGGG 1140  
AAGGAGCA GTACAGAG CAGAGCGTG CAGCCTCTCT TCCCTTTCA TTGAGAGGG 1200  
45 TGGTCCAGG AGCTGCGCG CTAACTCTG CTGAGACATA AGTGACTTT GCTTTTGGG 1260  
CTTATCTG ATNATGCTG GAGCCTTGC TCTCACTGC TAGATGAA CTTGAAATCT 1320  
TCACTACT CTTAGTCTT CAGTTTCTAC GTGTGAGAG CAGCTTTG GGCAGTCT 1380  
50 CTGTACAT CTGTAGACT TAAAAATTA TTCCAGGTT CCGTGAATA CCAATGCCAG 1440  
GGTTCCTAT ATCTGAGTT TCACTTGA TTATGACTG TTTTGGTAC CTGAATTTT 1500  
55 ATGTTTAC CTTAATATA GTCTGGCTG ATCATGAGA ATCTTTCTG GTCAACAGAT 1560  
CATTAATCT TATCAGGAG TTCTATCAG GCATCATGT CAGTGTGCT ATGCTGTTA 1620  
CAACTGGA TTTTGAAT AAAAATTTG TCAAAAAA AAAAAAAA 1669

## (2) INFORMATION FOR SEQ ID NO: 134:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1565 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 134:

5 CACTTTTGGT ATATACCTA AGTGATACC CTCCTTTAGT TACTGGCAA ACTGTGGCT 60  
15 TGGTTTATAT TGCATTTAC ACGATTACA AGCTGTAATG GTGTCTTTT TCTCTTGTG 120  
ACGGATGTTG TAAATCAGAG TATATACAT GTGTGCTGTT CTGTCTCTG GAGTTTCAG 180  
AGGATTACA CATGGCTTC AGTGTCTGTT ATAGATCTGC CTACCTTTGT GAATTCATCT 240  
GTTTACCCCT CTTCCTTTCA GAGAGACGG CGGATGTTG TTACTCTCTT GTCTTTTCTC 300  
TCTCTCTAC TGGTTATTTCT TGAATTAAAG AGAGATGCT CAGCTCGGTT GCTTTATCAT 360  
25 GAATATGTTG TGTGACCTTG CAGTCTCTCC AGATTTCAGC AACCAATGTC TACTTCTCT 420  
GACCAAAAT TAAAGAGAA AAGACAGATT TTTAAAGCA TCCATCTTTT AACAGCCGAA 480  
ACGGATGCT CTATGTTGCT GCACCTTCT GTTGTACTTC TGAATTCGAA CGTGTGTGAA 540  
30 GATCATTTTC TGAATTAAAC GTGATATGCT CAGAGTACC CTTCCTGTTG TTTTGTTAGC 600  
ATTGAATGCG AGCATATTTA TTTGGATAT ATACAGAGT GTTTTTCGAC TGTATTTCTT 660  
35 TTGCAAAGT TGAAGATGTC TTCTCTTACC TTTTGGAAA TAAATGATAT TGCATATTTG 720  
ATTCTCAAG ACTTGATAT GTGTAGCTTA TTAAGCTAG AATTTGATTT CATCTTTTCA 780  
TGAATGTGCG CTGAGTTCCC CAGCCCCCTCT COTCTTTT TTATGATGAG ATTATGACA 840  
40 CTCTCACTTA TTTAAACATG CAACATTTCT TGAATATGTA TTTTGAGGCC ATCTGAGCTC 900  
ATAGCTGATT CAGTAAACAG TTTCTATGCTG TGTATTTGAC ACTCACTACT TAAATCTGCC 960  
45 ATGCTGAAA TGTGAGGAA AATGATATCC ATGTGTGCTT GGGAGCATA TACATTTGTA 1020  
CATTTTITTA TACTGTGATT CTGTACATAT TGTGAGTTT GTTTGTGTTT ACGAGAAAA 1080  
50 AAAAAAGT GATTAACAA TCGAGAGCC AAGAGGTTTA CTATGATGC TTAGGCTGCT 1140  
CTGACCTTGG CTGGCCATA GACTTAAGC GCCAATTTAA TTTACGAGAG TAAATATTTT 1200  
TCAAAAGCCA ATTTTITTTT TGTATTTCT GTATGAACT GCGATATCA TGAATAGAAA 1260  
55 GGGAGAGCAA TAAAGAGAAA AGAAGTGTAT GTTCTGTTAT GTTCATGTAA ACTTAAGAAA 1320  
ACAGTGTGGA GCGAGGGCG ATCAGCGAA CTCTAGGGAC TTGTGTGTTG TTGAGAGCCA 1380  
60 TCCATACCTG CATTTTGCAT TCTTGATG TAAATCATTT GCCAAGACA AACTATTTCA 1440

TCATTATTTG TAAATACAC TTTTCCCGAG ACCTACATA AAGTTTCTGT CATGTATTTT 1500  
CTTCCAGTTG CATTAATAT TACTGAGTTG CATCATTTGA AAAAAAANA AAAAAAANA 1560  
5 CTGCA 1565

## (2) INFORMATION FOR SEQ ID NO: 135:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2007 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 135:

20 TCTAAAGCC CCTTATACC CGACTTTGTT CAGCAAGAT CCCCCTGAG GTACAGGCT 60  
GATTGTGCG CAGCTGAGC AATTCCTGA GGCACACTT GGTTCAGTT CAGATTTGAA 120  
25 GCTGTGTTG TTTTGGACC AGCAGAGCC AAACCTCCAG CCAACACAA GACTGTATG 180  
AGGACTCTCA GCTAGTGGC CTGTGAGAC CCCCAGGCTT TGTCAATAGA GTGTGTTGAG 240  
CTTCCGAAA GTACAGAGTG ATTGATTTG GGGAGAGCTG AATATTCACA GTTAAATGCT 300  
30 GAGCATATCC TGAATTTTAC TTTCTTATG CTTCGCTGCG AAGTTCTCTC TCTCATACAC 360  
ACACACGCC TTGCTCAGAA ATACAGAGC ACTCCATGCG CTCAGCTAT GGGACAGCT 420  
35 GCATTTGGCG TGCCTTTCTG TTTGCTTAG GAACTGCTGT GCTTCTTTG GTCTCACTCG 480  
CGAGGCACT CGGAGGTTG GACTTCGATT GGGCTGACAG CAGCTCTGCG ACGGCAGAG 540  
CGGGGCGTC TGAATCAGTC GTTAAACAA CACCTCTCTC TTGGCTCTCT GGCAGTTCTT 600  
40 TCTGCGAATA GTCTCTTCCC TGGCAGTTG AATGGGGGAA GCTGCTGCA CAGAGAGAG 660  
AGGCAATCCC GCTGAGGCT TAGGAAATG CTGAGAGCGG CTCCAGCAG AATATTCAT 720  
45 GGGAGGTTT TCAAGATGAA AATCATTTCT GCTGTGTTG GGGGCAAGT GTGTACACA 780  
AGCATCTCAA AGTCAMAGC CATCTGGGCG TGCCTCTCTT CTTTCTCAG CTCTGGGAA 840  
AGGATCTCCC CTCTCTCTCT ACTGATTTCC AGTGTGTTG GATTTGTTG GAGCACTGCG 900  
50 ACTTTTITTC TCTTTTCTT GATGAGCAA CATGCAAT GCATCTGCC CATTTTACTT 960  
TCAGTGATAT TTCTTTCTT GATGAGCAAT CTTTGTGCG CCTTTAGGAA GGAAGAGAT 1020  
ACACCTAGCA TGTCCAGGCG ACTGTTTAG GCGCTTTTAT ATAGTCTCT GTTAGAGTGA 1080  
55 GACTAAGGGA TGAAGACATC TCTTTATAAA AGCCCCCTTA GTAAATGATA AACAGAAACA 1140  
CTTAGAGTGA AGAAGGCTG TCTTCAAGAT CCAAGGTAG ATTGCTCTCA GTCTGATTTT 1200  
60 TTTCTGAG GACTTATCCC CTAAATATTT CTCGACTGCC ATACTCTCTC TTCTACCCA 1260

5  
10  
15  
20  
25

CGATGCTGCT CGGTGCACTC CTCAGATGCT CAGAGGGGTA ACCGATGCTC TTAGAGATTT 1320  
TGGGACCA TAGAATATCT GATGTGTGAA TTCTCTTTAA AAAACTTTAG GAGTCTTTGC 1380  
TACCTCTGCG TTGTGTAGTT GTTTGCTGCT TCAATTTAAA AGCGAGCTC TCACATTTTA 1440  
TTAGAGGTT GGGCTGTGTC TGTGCGATG TGTGTATACA TTTCGAGGG TGGCTGTGTC 1500  
CTGTAGCTTT TTAAAGGAA ACCGATGCTT CCGACTATGA ATCTGGCTATC TTCTATGCT 1560  
TCTAGTGTTT TGGGATACA TCAACGAGG GGTTTAAATTT ATCGAATGCT TGAAGCAATG 1620  
TTGAGAGGG GCTGATCAA ATTTGAGAG GGTATATGGA AGGGAGGGG GAGAGGAAAT 1680  
TGACATTTAT TTATTTATTT ATTTTAAATG TTATACATCTT CTTTATGTG TATCGAGCTT 1740  
GAATAGAAAC TCAATGCAAT AAATTAATCT GTTCTGCTCT CTCTCTCTGC TTCTCTTTTT 1800  
TTTTTTTTTA AATTAGGAT AACCAATTTT TGTTTCTAAA GTGATTTGTC ATTTGTGCTG 1860  
TATTAATCT ATTAAGGTT CTGTTTTAA AGGTGCAATTT TCAATTCCTCT GGGGCAATG 1920  
GTGCGAGA CATCTACAT GTAGAGAAC ACAGTGCAG ATCTGTGCTCT GATCTGAAA 1980  
AATTATTTTC TCTATATGAT TAAAGT 2007

30  
(2) INFORMATION FOR SEQ ID NO: 136:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1291 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 136:

40  
45  
50  
55  
60

CTTTTAAACC TCCGCTCTCA CACACATACA TATCAGGTG TTTTCTAGTT AAAAAGCCAA 60  
GTAGCTGAGA TTCTACTTTA ATGTCAATGC AGATTGCAAT TGATCAATGC CATTTGTTTT 120  
TTTCTCAATTT TTATGCTGTT GGGTCTTAGT TTTTAAATG ATTTAAAGAA CTCAGCAATG 180  
GTTTATTTTT CTACTGTATAC TTAGGGTTTA GGAACACTA CCGACTAGTTA TCTTTTATTC 240  
AACTCAATG GTCTACTGAA ACAAATGTC TTACTTTTCA TTATGGCAAT ATTTAGAGTT 300  
ATAGTAGTGG TTTCAGAAA ACGTTCTTC ACGATTGAC TTCCCAATCA ATTCATGTGA 360  
TCAATAGATT ATTCCATGCA AAGCGAAT GTTTGTITCA AAATTAATCT AGTTTCTGTT 420  
ACATTTAAAT TTGAGAGGT GAGCAATGCG TCTTTTTCAG TCTCTCTTCA TGTGATTTTT 480  
CTGATAGACC ACTATATGCA AACGATATCT GTCACTTACC AAATGTGTAA AATTTCCTGT 540  
ATTTCATCTT GTCTATATG TTAAATGTGA ACTTAAATCT TTGGCAATCT TGGCAATCT 600

5  
10  
15  
20  
25

CGTAGGCTG TTTTAAAGC TATGTGTAT TCTTACTAAT GTTCTATACA AGAATGCAAT 660  
TGTATATAT GCTGTCTAAT TCTAATGTT CAAATCAAT TTTGAGGTTC TATCTTATTT 720  
TATATAGAAA CAGACTCTTC AAAAATCTT CAGAGCAGC TTATATTTTA AATATCGAAA 780  
TATTTAAATA AACCGGTGG GTTAGATTAC TCATCTGTCC ACCAATGGG ACATTTGCAAT 840  
GCACTGGGGG CTTAAAGAC TTAGAAAGAA CTTGTAGTA AATCTGAAA ATGAGCAAT 900  
CCCACTTTGA ATGTTACTG GAGTAAACC ACCTTTACCA CCCCATTTAC AGCAGCGAG 960  
GCCATTAAC CAACTTGGCT CTGATCAAT TTTCTTTTCT TCATTTGTGA TGTCTAGAT 1020  
CAAAATGCT GTTCTACAT GTTACAGGCT TCTCTTTTGT TTCAATTAAG ATTTTATGCT 1080  
TACTTTGTA TGGACCAAT AGAATTTCA CAGACAAA TAGAGAAAT TGTGTTTAGA 1140  
TATTTTTCAG AAGTCAGAG ATTGTGCA ATCAATTTAT TTGCTTTTT AAAAATTCAT 1200  
TTAGCAATTT CAGAGATGAG ACTACTCAGA AAATTAATTC ACGTAATTTCT CTAGAGGTC 1260  
AATATTTTT AATCAATAT GATCAATA A 1291

(2) INFORMATION FOR SEQ ID NO: 137:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1906 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 137:

40  
45  
50  
55  
60

GGCAGAGGA CTTACTTTTG TTACAGACA TGGTTGTGTC CAAGGTAAAA CCAAGTGAAT 60  
ATTTTGTGAT GCTTGTGTCG CATCTGTGAC TTGTTTTGTC AGTATCAATTA TTGAGACTTC 120  
AAATGTGAA TCTTTTAAAC ATCTGTATTA TTGTGTGTC AGAGCTGTTT ATTTCAAAAT 180  
GTATGAAAT TCACTCTAGT TCTGTGATA AGATCAATCA GTTTTGAAG GTTACTGAT 240  
TTCTCTTCC CTCTTAGTTT TTACCAAT ATATGAGAA GAGTAATGCT CAATCTTAAAC 300  
ATTTTGTGTT AATGTTTTAA TTAAAGCTT GGGCAGTGT CAGCAATTC TACCTAGTGT 360  
CATTAAGCA AAATACTTAC ATAGCTTTCT TAAATATGAG GAATGCAAT ACATTTTATG 420  
GAGAAATTA GTTCTTGTCC ACGCTACTT TAAATGCTTT CCAATATGTC TCAATCAAAAC 480  
TTTGAATAT GGAATCTTAC TATTTGATA GAATGTGTA TGTATATAT ACATACATAC 540  
ATAGCAAT ATGTGTGCT GTTGTGTAT ATATATAT ATGCATGCTG TGAATCTTGA 600  
CTACACACA TAAATCACTT TTTAAATTC AGCAAGGCT AGTCTGACAC GGTGATTTATC 660  
CTTTGAGGC TGATTCGCTT ATTAATGCTT TATTTAGCTT TTACTCCGAG TAGCAAGGGA 720

5 TTCTAAGTTA GTTGCACTTA CATGATTATT GTGATTTAAA ACTAAGAAATA AAGGCTGGAT 780  
TTTCAAGAT AAATTGAAAT TCGTGTGGT GAATATACAA CCAAAATACT GAATCTGATG 840  
TACATACAGG TTCTACAGG AAGATGGT ATATTTTACA ATTGGAGAT TTAAATACCA 900  
GGGTACCA GAAAAAGTGA CTTCATACCA TGGTACAAAT AAGTAAGGGA TCGTCTCTCG 960  
10 GTTTCCTTTT GGCCTTTTCA AGATTTTTAC TTCTCAGGTT ATTAATTCAAA ATTAATGAT 1020  
AAGTAAGCA ATAGAATTTT TAGGTAAGAA CAACAGATGG GGGTTTCTG GAGTGTTTAA 1080  
TGTCAAGGC ATTTTATGTA GATAGAGCC TTGTCTCTGC ATTTGATGAT TTGATTAATT 1140  
15 TTGTCTTAC AGTTAAATCTT CCTCCCCAA GTTTGCTATT CAATCAACT GCGTGAATGA 1200  
CAATCTAGT AGTCTGATCT ATTTTCTGA GGAATAGTTT GTGATTCOA TCGAGGTGTC 1260  
20 TTCAATACCA TTACCTCTAC ACTGACAGAG AAGCAAACT CATTATTTAG AATTAAGTCA 1320  
CATGCTATG GGGAAATAG TTCTGAAGG CTGGAATGAT AAGATGAGC AAAAGTTGGT 1380  
25 CAGCTGGCT ATGAGTGGT GGCATTAATC TCTAAACATT CCAAAAGACC ATGAGCTGAA 1440  
CCTAAACTCT CTGAGGATC TGGACAAAG GATATTCAAA ATTGCATTT GAAACTGAC 1500  
CAGCTAATCT GAGCTTCAGA GATAGATAG CCAATGGGCC AAGCCATTT CAAGTACAGA 1560  
30 AATTAAGAG ACTACAGCTA AATTAATTTG AATTAATTTG AATTAATTTG AATTAATTTG 1620  
CTTAATAGC ATATTTGTAA ACTCAGACT GAGCAGAGT GACTTTTACT TTCTAAGTTT 1680  
35 GATTAAGT TGAATGTTCT CTATGCTGC ACCCTTCCC TTCCCTTTC TAAAGCAATA 1740  
GTGCAGACT TAAATTTT TTGCTTCAGA ATTTGAATGA AAAATTAAT GCAATGATTT 1800  
TTTTCTTTT GCAAGACACC TGTATTATCT CTTCATTTAA TGTAAATGTC CCTTTATGCT 1860  
40 TTGAAATTA ATTCTCTTT GTAAAAAAA AAAAAAAAAA AAAAAA 1906

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1935 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 138:

55 TCTGAAGTAA TCGTACAGA TCCCTCTGAG GATTCCTTGA TGGGCTGAGC AGCTGGCTGG 60  
AGCTACTACT GACTGATGAT CATCTGATG AAGGAGAGCTT TCTGATACG GATTCATAGC 120  
TCTATGTTTT ATATATATTT TCACTGTAC TTGCACTGCA CTTTACAGA GAGGAAGTAA 180

240 TGAAGATTA GCTGATGCG TCAAGGTGAC TTAGGTAAAT TGGCAAGTCC ATGCTTCCCA 300  
CTGAGTCTCT CAGGTACAGA AGCTACTATC TGTGCTAATT TTGTATCTCT TCTTTAATAT 360  
5 GTGCTAGCT TTGCAAGTCT TAGAATGGAT CCTGTGTCY TTTTACTTTT GAAGAAATCA 420  
GTTCCTGCT CTTTTCGAA AAGAAAGCAA AGTGCATATG TTTTATCTG GAAATTTACC 480  
10 CAATAGATG AGGTGAAAGC GATCTAGATT AAGGCTCTCT GTAAACAGAA AATCATATCA 540  
AAACATATCT TTCCCATCTG TTTCCTCAATG CCTGCTACTT CTTCAGATA TTTCATTTCA 600  
GGAAGCAGC AGTTAAAGCC GTGATTTTGT TAGTTAGGAA CTGGGATCA AAGCCTCTTC 660  
15 CACTAATGG CTATGCTCT GAAAGATTTT TTTTATTTT TTTTATTTAA AGCCTTTCTG 720  
AATTTTCACT TTCTATGCT ACTTCAGAA ATTTGTTGCA GCGTTGAGAT AATGCAATTT 780  
20 TAAAGGCTCT GCGATTAAG AAGTCTCTAG TTATGAAATT AAGGTTGCT TAAAGCTGTA 840  
AAGTCTTAA ACTTACAGT AATCAAGATG CATTTACCC CACTGACTTG GACATAGTGT 900  
AAACTAGCC AGAGTCTCT TTTCATTAAT ACTTACAGT TATTCATAT TAAATTTTGT 960  
25 TATTCATTA TCTTATTTAT CTAACTAAA GCTTCTCTTT TATACACT CTCTTATTT 1020  
TGGATTAAGA TAAATGACA CATTACTTTA ATTTCTAGT GGTGCTCTCT GATGCTTCAAT 1080  
30 TGTAGTTAAG GACATTTTCT YTTTTCAGC AGCTGTGTAG GTCCAGAGCC TCTGAGAGAG 1140  
GAGGCGGTTA GCAATGAGCC AGCAGGAGC TGAAGTGGGA AACTCAAGGT TCTTTTACT 1200  
GTGCGTATG GAGCTGCTTT TCTGTGATCG GTTTCCTTAG GCAATGCTCT GTTCCCTCC 1260  
35 TTGCTATTC CAGCTAATTA CAAGCTGCCC AAGCCATGTA GCGTGAATCT ATATATGATC 1320  
AGTCTGTGT CTGACTCTCA ATAGCCCCAC CCAAGCTGCG TATAGATTTA CAGATACATT 1380  
AATTAAGCAA CTTAAATAT TGAATCTGGT GTTGTGTGTA CATATGCTTA TGGCCAGAAC 1440  
TGAAGCTTAG AGTTAATAT CATATTTTAG GTTCTCCAG AAGGACAGAA TTATGATGAT 1500  
45 ATATGTATAT ATGAAAGGGA GTTTATTTAG GAGAACTGCG TCCGAGTT AAGAAAGGCAA 1560  
GTGCGCAAT AAGCCTCTG CAAGCTGGT TAAAGAGAG CCAATAGTGG CTACGCTCA 1620  
GTTCMAAAC CTCAAACTG GGAAGCTGA CAGTGCAGC AGCTTCTAGT CTGTGCGCA 1680  
50 AAGCCAGAG CCGCTGCGA CCAAGCCACT GGTGCAGTC CTAGATTTCA AAGCCTGAG 1740  
AACTGAGT CTGATGTGCA AAGGAGGAA GAGTGGAGA AAGCCAGAG ACTGAGCAA 1800  
CAGGTAGAC AGTGTCTTAC ACCATATG CCAATACAAA GAGCTTACCG ATTCCTTCT 1860  
55 GCTTACTGGA TCCCTGAGT TGCCTGCTCT TCTGACTCT CTAACTCTT TCTTTAAGAG 1920  
CTTTCTATTA CATAGCTGT CTAAAGCCC TCAATTAAT TCTCATGTTA AGTTTCMAA 1935  
AAAAAAAAA AAAA

TTCTAA

1446

(2) INFORMATION FOR SEQ ID NO: 139:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1446 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 139:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

NCGCCCTTG GCACAGTCA GATGAGCAG GTTCTGCCGG GAGGCCCTC AACTTCCAGA 60  
GAGGACAGAC ACGATTTC TCCTGGGGGA GGGAGGATC CAGCATCTT GATCTGCTT 120  
GGAGCTTAT TTTCCTGTTG CAGATGCA TTCTCTGAG TGGAAAGAG TTCTTGCATG 180  
TGGATGTG TTTCCTCAGG CAGAGCGGCC CTCTTTTCCC AGCACTTCCC TGCCTGCCCC 240  
AGGCTCAGG CCAGCACCA GTTCTCTCTC ACATGCCAG TGGACAGAG CTCTTAGTTG 300  
GCAGAGCTG AGAGGGTGA ACAAGCCCTG AGGGAGCTC GGCCTTTGCT CCGAGATTGG 360  
GGGAGGGGG TTGTGCAGG TGCCTCCCGG AGAGGCCAG CATGTTTAC CAAAGCCCTC 420  
ATTGTGCTC GAGGAGACC TTATTCCCGG GCTTCNAGC ATTGCTCAT CTCTCCAAAC 480  
TGGTAGTGT GATTGACGG GAAGACTCC CAAATGTGC CAGGATTTC CCGTCCCGG 540  
GCAGGCGAG GGAATTAAG GCGAGCGAG ATACAGGGG AGGGATTGTG CAGGTGAGGG 600  
GCTCTCCGC TTCTCCCTCT CTCTCAGCA TGTCTGCCCC ACCGTGCTC AGTTCTCGCT 660  
TCCCTCTAT CTCTGTCCC CTCTTTGAG CTGTCCCAT CTGAGTGA GACAGACCTT 720  
CTCTCTACT GACACCTTC CTCTGACCA GCGCCCTTCC TTGTCTGAA AAGGAGCTT 780  
TGAATGTGT AGCGAGGAG TGGGAGAAA GGTCTCACC GACAGGTTGG GAGATGAGG 840  
TCAGCGTTGC TGGGAAACG ATGAGGGGG CAGTGGGAC AGGCTTTGG CAGACACGAG 900  
CAGGATTAAT TTGAATGTG TGAAGTACT CCGCGAGGC CTTCGGCTTG GCGATTGCG 960  
AAGAATGA TTCTCTGAG GGTTAAGGG ACACATGGA CAGGGGAGA GTCTCTATCT 1020  
GCTGGCATTT TGTGGGTTT TACTGCGAA CTGAAATAG GGTCTGGGTT GGTCTCTTCA 1080  
CTGACACCA AATCAGAT CCGTGTCTT GATCTCCGAG AACTTTGCT CTTCAGTTC 1140  
CTTCTCTTC CTACTTCAAT CGATGAAA TTAGTTATT TTGATCTTT TCCCTTGCTT 1200  
GGTCTAGCT CTCTCCAAAC AGCATGCC TCCAAATGCT AGAGCTTG GCGCTGAGC 1260  
CTGTAGACG ATGCGCTCAG AATTGGGCA TGGAGGGGG GGTGGGGAG CCGATGATC 1320  
AGCGAGGAC TCCANTGCC AGCTCTCTC CCGAACA TCCCGAAT CCGTTATCCC 1380  
TACCCCAAC CTTTGGGCT CTGTACAT TTTTAACTT GCGAAAGT GAAGAGATA 1440

5

(2) INFORMATION FOR SEQ ID NO: 140:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1109 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 140:

15  
20  
25  
30  
35  
40  
45  
50  
55  
60

TTTTTTTTT TTGATATCA AATTGTCTTT CTCATATCA GAATTAAGCT AGGAAACAC 60  
TAACCCAAA ACTTCTCTA GAGCTGTTC TTTCGAGCA GCATCACTTA TTGCGCATTA 120  
AGACTAGTA TAAAGGACC AGCATCCCTA CTTCGGTCAAT GGGATTAAAT TTATAGCAT 180  
TCCATTTTC TAGTCCACA TGTCAANTG GATTGTGATG ATCTTAATCT ATATCTTAC 240  
CTTATATTA AAGATCAAA GATATATCT CTATGACAG ATTGAGATA GGAATGAAA 300  
AGTTCGAGG ATGCTTTAT TCTAATGTGA GGGTAGGGA AATGTGATA AATTACTGG 360  
GGTAGGAG GATGTGCT TTAGTTGAG TTCTCATTTT TATCTCCAG TACTGACTTG 420  
TGGGAAAC ATACTTTTC ACTGCCAGT ACTGATGCA GAGGCTCAAT GAGATATATA 480  
TGTGGAGT GATGCATTT CATTATATG CAACATAGC TCGATTAGA CAAGTTGTT 540  
GTTTGGAAA GGGTTAAG CTTTAAGTGA ACAATCTAG CTATGCTGA ATGATAGAG 600  
TATATTAAT TCAATTTTT TATTTCTCT TTGTTAAGG TCCCCATAC TTCTCTGTC 660  
GGAGATGA GAATATGAT TACTTCACTG TTAGTTTCT TAATTTTTT TTTCCTCTAT 720  
TTCTCCCTG TACTTTGTT GGAAGCTAGA AATCTGTGG TTATATATAG GCGAGCTCTT 780  
TGTGAAGTG GTTATATCA CTGGGAAG GCGATTGAA ATCATTAGA ACCAATGAT 840  
TTCTTCCCC AGGAACT ATTCCTATTA GATAGCTGA AGAGCTCT GTGAGAGCT 900  
CAGCTCCAA CAGAGTCA GCAGCTCTA TAGGATTC CATGATTAAT GACTTCTCAT 960  
TCTGTTTAT CAGAGTCAAT ATATGTCTA CTTCAGAAA AATTAAACAG TCAATTAGA 1020  
AAGAACTCA ATCTATTC TAGCATTTT AATAAAGT TAAACAAA AATTAAAGG 1080  
GACACTGAG GGGGGGCGG AAACCAAT 1109

(2) INFORMATION FOR SEQ ID NO: 141:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 497 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 141:

5 TAGGACTAAC TTAATTCCTT TTAATCACTT TTAATTAAT AAAAAATTTT ATTCTTTCA 60  
10 ATTTTCCTCT AATTCCTTA RCTCTCTTA TAAATGTGTA TATCAATG ACCCAACT 120  
CAATATCCTT AACATTTACT CTCAAAAGC TTATTATTTT TATTTTTTT AGGTAGTTT 180  
15 TTCTGTGTCT ACTCTGTAC ATGATTTTC TTTCAMATCA TTCTGTGTCC CCAATAGAAA 240  
ATGCCCTTTTA TTTTGAAGA TCGTGACTT TTTAGTATG CATGAGTGTG CTAAAGGCCA 300  
GATATCTTTT CACATTCCT GTTGCTTTTG AACTTACTT TTTATCTTC CATCTTTACT 360  
20 TTAACCCCTG AAGTGCAGT CTTCACTCAG GSCCAGACCC GGCCTGAGGC CTTTGTGGA 420  
GATCTGCAC CACACGAGA AGGCTGACAC CTGGTTACTT GTACTCTTTC ACTGTATATA 480  
AAAAATTTA TCTAAAA 497

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 269 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 142:

35 ATGAGGGAGA GCGAAGTGC CTGCCAACC CTCCTCTCAA GGAATGCGCT TCCCGAGAA 60  
40 TCCCGACCAAC ACATACCGTC TTCTTTTTT CTAGTCAAC TCTTCTTTAT TCTTGGCTT 120  
GCTCCCTCC TTTCCTGCC TCTCAACTT TTACTCTGAG TTTCTATTTT ATGGGATTTG 180  
45 GGTGTGAGT TAACTTACA ACGTGTCCC CACACGAG TCTTCCAGA AAAAATACA 240  
AAGAAATTA AAAAAAAA 269

(2) INFORMATION FOR SEQ ID NO: 143:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1269 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 143:

5 TTGATTAAGT ATGCTCTCTC CGGTACACAG GAAGAGTCTG CCAAAATGAA GGCATGAC 60  
TTATCACTCT CCACAGCAAT CTGTGCTCTG CTGTGCTCTG GCTTGGGCTT CTTCGGCTC 120  
TTCTGGCTGC TCAATGCTGT GCGCGGAGAG GCTTACTTGC GAATGCTGT GTGTGTATTC 180  
ACAGGCTGCA CTTCAAGGCT GSCAAAGA TGTGCNAAG TCTTCTATGC TCGCGGTGCT 240  
10 AAATGTGTGC TCTGTGCTCG GAATGTGTG GCTCTAGAG AGCTCATCAG AAGACTGACC 300  
GCTTCTCATG CCACACAGGT CGAGACAC AGGCTTACT TGTGACCTT CGACCTGACA 360  
GACTCTGGBB CCAATATTTG AGCACAGCT GAGATCTCTG AATGCTTTTG CTATGTGAC 420  
ATACTGTCTA ACATCTCTG GATCAGCTAC CTTGTATCA TCTGTGAC CACAGTGTAT 480  
GTGACAGAA GGTTCATGGA GACAACTAC TTTGCTCCAG TTGCTCTTAC GAAGCACTC 540  
20 CTGCTCTCA TATCAGAG AGGCAAGGC CAGATTTG CACATGACAG CATCCAGGC 600  
AAGATGACA TTCTTTTTC ATCAGCATAT CGAGCTTCA AGCAGGAC CCAAGCTTTC 660  
TTTGACTCTC TCGTGTGGA GATGAGACAG TATGAAATTT AGGTGACCT CATCAGGCC 720  
GGTCACTTCC AACACACT CTCTGTAAAT GCAATCACT GCAATGATC TAGGTATGGA 780  
GTATGAGCA CCACACAGC CAGGCGGGA AGGCTGTGAG AGGTGCGCA GATGTGCTT 840  
30 GCTCTGTG GGAAGAGAA GAATGATGT ATCTGTGCTG ACTTACTGCC TTCTTGTGCT 900  
GTTATCTTC GACTCTGC TCTGTGCTC TTCTGAGCC TCTGTGCTC AGGCGCGAA 960  
AAGAGCGAA ATCAGAAC ATCTTACT CTGACAGCC AGGCGCGAG CAGAGAGCA 1020  
35 GCACTCTTAC GCTTCTTAC TCTACAGGG AAGTGTGAT TTGTGAGAC TTTATGAG 1080  
ATTTGTCTCA CAGTGTGGA AACTGAGAA AACATCTC GTGCAATCT GCTGCGAGAG 1140  
40 GACATCAAA AAGCAGACA AGCTTCTTC CAGGTGAGG GCAATCACTT AAGCATAAA 1200  
TATGAGCTG GGTTTTACA CTAAAACTA GAATTAACA TCTCAACAG TAAAAAAA 1260  
AAAAAAAC 1269

(2) INFORMATION FOR SEQ ID NO: 144:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1946 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 144:

50 AAGAGCAAA CTATAGGATA ACACAGGCC CTTTGTGAA ATAAATGCG ATTGAGTGT 60



5 TTATACCTCTT AGCTGTGTTTA CTATAGATGT AACATATGCT GCTTACCCAC CTCAMAMTGT 120  
CTGTACTGCA AGAGGGCCCT GGGCTCTGCG TTTCCTATAT CAGCTTTTGG CAGATTTGTA 180  
GTCCCAAGAA AGAGCATGCG TGGCAGATCG TACGGATGTT AACTGGCTTG TCGATGCGCG 240  
ATGGAGGACA AGGGGTGACA GATGGCTCCG TGGCTTACCG TGGCAGTACG GAGACAGTCC 300  
AGAGATGCT CTCTGTGCA CCGGGTGTG TTGTCTCTCG TGGTCTGCA TGTCTGTGG 360  
TCACCTTTAT TCTTGAACT GAGCTTTACC TGGATCTGCG TACTGAGGCT AGAGCCACA 420  
GAGATGCG GTTGGGCTTG TGGGCCCCAA ACTAGGGGGT GTGGTTTAT CAGATGTTG 480  
CCTTTGTCT CTATAGATA GGAATCTACT TTTCAGGGA ATTGTCTC CTAAATATAT 540  
TTGCTTTACC TTGCTCTTTT CTGTGTGCC ATATATCAG TGTATAGCT CTGAGCAGGG 600  
TCACATTTGG CCMAAGTGA CACTGTCTTG CTGATCTC CTTTGGCAA CATCAGGCTC 660  
AGATTTAGG ATAGGCTTC CTAGGGGACT GACTTTCTG GATGGGGCG TGTGTGTGA 720  
CAATATATTT TCAATTTACC TGGAGATGTT CAGAGGCTG CTCTGAGGCT GAGTGTGTT 780  
CCCCCTTGG TGTGTTCAG TGTGAGGG ATACATCTCT AGGCTCTGCG AATTCAGGCG 840  
CAGAGACTC CTGTGTTTGT GTTCCAGAT CCTGTACTAA GAGGGTCTG GCCAGAGGAA 900  
CAGACAGCT TTTCACATG GAGGCGAG GAGACAGTG GTTTGCTTGG TGTCTTACT 960  
GTCTGAGCC TGGTCTCTTG GGGCATTTGA AGGTATGATC TGTGATCTCT GGGTTTTTG 1020  
TGGCTTTTGT CAATGCTTC ACTTGAAGG AGGAGAGCA GTCTATACTC TCCGAGGCT 1080  
GCTTGAAGCT CAATGAGG TATATGAG ATCTGTGAT ATTGTGATG AATATCTTTG 1140  
GTATTTATAC ATTTCAGCTG CAGTCCACA ATTTCGAGG TACATCTTAA GCTATTAAC 1200  
ATCATTTCTT AAGACAGAG GATAGCTCTG ACTCATGGA TCAATGAGTC CATGGCTGTT 1260  
TCCAGTTCC CTTTTCTCTT CTTCAAGTTT TGTCTCTTCC TGTGTCTCC CCAGCAGGG 1320  
AGAGACTGCG GGTGTGATG GAGAACAGAA TTAGAGATAT AGCAATTTGA CCCAGATGCG 1380  
AACAGTGGCG AGCTACTGTT GATGAGGAG ATTAAGTCTCT GAGGAGCTG GAGTCTAGGT 1440  
GTGAATGCTG TATTTGACA GGGATTTAT ATCTTGGGTT CTGAGAGCTTT CACTCTCTCG 1500  
TCATGCTCTT CCTGTATGAC TCCATTTGCA CAGGCTCTGA GTTGCAGCTC TGCACCTTAA 1560  
ATCACAGCTT GGGCATTTCT TGGGCTGCG GGTCTGCGG TTCTGTACTT GTGTCTAGCT 1620  
GTGGGCTCTG ATGCTGGAGC TGGAGGGTTT TGTGTGTCTA GACTGTAGCC TGTAGCTCTT 1680  
GGCTGTGTA GAGCCCCCTC CTGTGGCTTC ATGTGGCTTC GTTTTATTAC ATCATCAGGA 1740  
AGATGGAAA GTTCAGGCG AATTTTCTG CCTACAGAG GTTGGAGAG AAGGACACA 1800  
GTATTTTAT GAATTTACA TATATCTTGT TTTTCTTCA ACGAAATGT TATTTGAGCG 1860

5 ANTTCTACT GCTCAAGTTT GATGTGTTTA TTCAAGTAA ACTGTAGTTT TCTGATTTA 1920  
AAAAAAAAAAAAAAAAA AAG 1944

10 (2) INFORMATION FOR SEQ ID NO: 145:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1021 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 145:  
TCGACCCAGC GTCCTCGGGT GGCAGCGCG GAGTTCCGCG TCGAGACCCG TGTCTGTGGC 60  
CGGGGCTTC ACCATGGCTT GGCAGAGCTT GCACTACACC ATTCAGATTC CGATCAGCC 120  
CTCTGAGCG CAGAGAACCA GCTCCAGCCC AGTTGGGAG GAGCAGAAA CTGCGAGCC 180  
TGTGTGATTT CTTTTGGGCT GGGTGTGCTG CAGGACAG AACCTTTCCA ATTCAGTGC 240  
CATCTACAC AAAAGGGGCT GCATCTATAT CCATACACA GCTCCCTGCG ACATGCTTT 300  
CTTCTCCAG TCACTGAGTA TCCCTTCACT TGTGTGTTTG GGCAGAGCG TGTCTGAGCT 360  
GCTCTTTGAT TATGAGATG AGAGAGGCC CTTCTCTTTC CATCTCTTCA GCACGCTGG 420  
GTCATGCTG TACCCCTAGC TGTGTAGGCT CTTCTAGACC GTCCTCTTCT GCGGCTGTGG 480  
TGTGTGGCG ACCATCTTTG ACAGGCTTCC TGTGTACAG AACCTGTGAG GGGCTCTGCG 540  
GGGCTTTGCA GCAATCTTGG AGCCCTGGCG CCCATGCTG GGCCTTTTGC TGTGTGTGG 600  
CTTTGCTCTG GTGGTGTGTC TTTTCACTT CTTCTTGGT CCATCAGAG GCTCTTTTCA 660  
CAACCACTTC TATGAGAGCG TACAGAGCC GGGCTTTCGG TGGCCCGAGC TCTACCTTTA 720  
CTCGAGGGCT GACGAATGAG TCTTGGCCAG ACACATAGAA GCAATGCTGG AGGACAGCTT 780  
GGCAGCCCGG GTCTGTGGCG GTTCTGTGGA TTCTGTGTCA TCTGTACAGC TCAAGCAGCT 840  
CCGTGACTAC CCTACTTACT ACACAGGCTT CTGTGTGGAC TTCTATGCCA ACTGTGTGCG 900  
CTGTGAGCG CATTCTCTCA TCTCACTCTT GCTTCAGAAA TAAATGCTTG ACAGCTTCCC 960  
ACAAAAAAA AAAAAAAAAA ACTCGAGGGG GGGCCCGGTA CCCAATTGCG CCTATTAAAG 1020

(2) INFORMATION FOR SEQ ID NO: 146:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1285 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 146:  
GCGACGAGG GCGCGACGGC AGCCATCGCG CTTCACGTT CCGTCTCTCG GTGACGCGC 60  
AACCCGAGT AGCGATTCG GTTCCCTCCA GTCCGAGCC CTATCAGAT TCGATATGTC 120  
CTTCATATTT GATTGCAATTT ACAGTGGTATT CACGAGTGT CTACAGTTTT TAGGATATATA 180  
TAAGGAAACT GGTAAACTCG TATTCTCTCG ATTGATATAT GCGGAGAAA CAACTATGCT 240  
ACACATGCTTA AAGATATGACA GACTTGGACA ACATGTCCCA ACATTACATC CCACTTCCGA 300  
AGAACTGACC ATTCTGCGCA TGAAGTTTAC AACTTTTAT TCGGTTCGAC ATTCTCAGC 360  
TCGAGAGTG TCGAAAACCT ACCTTCTCTG TATCAATGCG ATTGTATTTT TCGTGGATGT 420  
TCGAGACAC GAAAGGCTGT TAGATCTAAA AGAAGAACTT GATTCTACTA TCACAGATGA 480  
AACCATGCT ANTGTGCTTA TACTGATGCT TCGGATATAG ATCCGACGAC CTGAGCCCAT 540  
CAGTGAGAG AGTTGCGAG AAGATTTTGG TTTATATGCT CAGACACAG GAAAGGGGAG 600  
TATATCTCTG AAGAACTGA ATGCGCGACC CTTAGAAGTT TTCAATGTGTA GTGTGCTCAA 660  
AAGCAGAGT TACGAGAGC GTTTCGCGTG GATGCGACAG TACATGATTT AACHAAACT 720  
CACATGCTT CAGGCTCTCA AGCTTCAGGC TTAACAGAG ATTGTGATTC TCAACATGCA 780  
TAACCTGAT TCATAGACT TTGCTGCTGT ATAAACGTA TGTTTTTTAT ATATATATA 840  
TTAAATCAAC TTAAATTTCA TCGAATTTGA AAATGATTC AAGTAAATTT GAGTATCA 900  
ATGTTAGCTT TCTAATTTCA TAAAGTACTT TGGTTTTTAC AGTTTATAT TCGACATCAC 960  
CCGAGCGCCA TTGTATAGA GCAACTTTTC AGCAATACAT TTGAGCACT TTTTAAACAC 1020  
ATGAACTAT AAACCATAT TAAAGCTCA TCAATTTAAA TTTTATATGT ACTTTTCTCG 1080  
AACTAGTTTT TAAATTTTAT ATTAATATGC CACTATCTT AAGTGTACAG TTAATATATA 1140  
GCTTATTCOA TGAATGATG ATGCGTTTCA GTTTTCATA ACTTTTTTTC TTATCCGAC 1200  
GTATGCAAT AAACAACT CTATGTTTGG GCAAAAAAAA AAAAAAAA ATCGAGGCGG 1260  
GCGCGTACC CAATCGCC TAAAG 1285

(2) INFORMATION FOR SEQ ID NO: 147:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1386 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

60

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 147:

5 GCGACGAGT GCGCGACGGG TCAGTGTGTC TCTCGGATCT CCGGACAGGT GAGCAGCTTG 60  
ATGAAAGCCA CCGTCTCTGAT GCGGACCTG GCGGATGCA GAGATCTGT GCGGCGCTCC 120  
CGAGGCGCG CCGAGACCGG TTACAGTGA TTTCTGATTT TACATGACC TTGAGCAGGT 180  
TTGCAATATA TCGAAGCGCA TCGCTCTTT CTACATATAT TCTGATAT TCGAAGATCA 240  
TCATGAGGA GTGTGCGAAA GAGCTGACAG CGTCTCTTCA CCAATTTAC CCAATTTAGA 300  
TCGACCCA CCGGACCTG AAGGAGAGC TACTCTATAT GTTGTATATG TCGACCGAAG 360  
CGCAATCT CTATGTGAG CAGAGATTC AGAATTTTCA GATAGCCGAG GTGTGTAGAG 420  
AGTCAATGC ATGTCTGAG GAGGATATA AGACTTTCTT CAGCAGACTC TACCTTACA 480  
ACATTTCTCT TTTCAATCTT TCTCGGCGCA TTGTGTATAT CTTGTAGAA ATTATCCGAC 540  
AGATGAAAT GTTCACGCC AACNTCCACA TGTGTCTTAA TCTATGAT TTTTATGAG 600  
ATGTTTCTT CCGGAGCTT AAGGGCGAG TGAATACAC ATACACAG AACAGCTCTG 660  
TGTCTGAAA CTGTGTTAC TTCCAGCAG TTGAGGCGCA AACCAATGTC ATCTCTGAG 720  
GAGCTCTAT CCGGAGCTC ACATTCGCG ATGCGTTCC TGTGTGACAG AACATTTCTA 780  
AAATGCGCT CTTGATGAC AGGTGCGAG AGCGCGGGA AGCTTACATG GACTCTATG 840  
ACATCTGCT GAGAGAGAC GAGACTCTTG ATGTGTCAA CCGGCTACTG CAGCAGATCC 900  
TGTCCAGGG GTTCACCTG GAGATGCAAG GCGCTCGAG GCGCAGGCTN CCGAGCGCC 960  
TCAGCGCTT GTTGAAGAG GCGGCTCTCC CAGATCTGC TCCCGCTGCA ACACAGAGCA 1020  
GAGCGCGCG TCGCAGCAG TCGCTGCTGC CTTCGCGCC CTTCTCTCTT CTTTCTCTG 1080  
AGCACTTCA TCACAGAGG CTTGAAGGAA CCGCGCATG TCGCAGGCGA CAGGCACTGT 1140  
TCTGTGCAA CTTTGACCA CAGCATCTCA GTGTCTAGG GATTGTAT TCGAGGAT 1200  
TTCTCAAA TTTTAAACA TGGAGATTC AACAATAT ATGTGTGAA ACAGATCAA 1260  
ATTTTAAA TCAAAAAA GCTGCTCTGA TTGAGGGAT GTGCGTGGG GTAGAGCTG 1320  
GACTCTTGG CTTGCGGCGCA CATGCGATGC TTTAGGAGC ACGTTTGG AACCAACAA 1380  
AAAAA 1386

(2) INFORMATION FOR SEQ ID NO: 148:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2098 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 148:

5 AACCCTCTTC CCCGCCCTTG GCACTCTGAC ATCTTAAAGC TGCAAGCTCG TCTCTTCTG 60  
TGGGTAAAGC CATCTCTCTG ATCGAAGATT CTTGCAACTG ACGAGGAGG GGGCTGTGAA 120  
CCCTAAATGG GTTAAATCTC CTCGACCGA GACTCTTAC CTTGAGATG GCACTTACTC 180  
CTGGCAATGG TTCTGAGCA TGGGCAAGC ATGTAAGGCC ACCCTACTGT GTTATCTGCG 240  
CCTTCAATGA ATGTAAGGCC TTGGGAGAA GGGCTCTGAG ATCAAGGAAT GCACTGAGC 300  
CTTGGCTTAG CGACTGGCT GGAATCTTCT GGTCTATCTG ACTGTGATTC CAGAGCAGA 360  
AACTGTATTT CTCAGATTC AGATCTTACC CAAGACCAAA GATGTATTTT CAGAGAGCA 420  
GACCTAGAAA TGGGCTCTTC TGGATTTCA GATCAAGCA AAGCAGCGAG GGCCTAGGAG 480  
CTTCTGTGGG TCTACAGAG AAGGTTCTTG TGAAGGCTAT CAGTGTGTTG CTCTTAGCTT 540  
GCTGTAACT TTGGGCTCTC GCGCAAGCC TGGCAACTC CCGTGGCTGT GATGGCTTTC 600  
TGTGGCATCC TGGCTTCTCC CGAGCTCTTG CAGATCTGCC TCCCTAACCA ACTTCTCTG 660  
GGCTTGGCT GTCCACTGCG CTGATCTAT GTTCAAAACA CTGAGCTGCG AGGCGAGCA 720  
CTTCTTCCA CTTCAAGAT AGTCTCTGCG CCGCTTCTCT TGGCATTAATA ACGACTTTTA 780  
ACGATGAGCC ATGTGCTTCT TTGGCTTCTT CTCTCTCTTT CCAATCTTCT GCTTCCAGT 840  
CACTTCTCTG GCACTATGCG ATCACTGTC CCGCACTGT GTGGCAACAC CATGTCTCTT 900  
GTCAATACAG AACTGCTCTT GAGCTCAGG CTGACACAG ATCAAGCCAA GCTTGTATCC 960  
TCAAGCCCA CTTTTCTTAC CATTCTCTCT CCGTCTCTCT TCTTTTCCA CAGCAAGCTT 1020  
ACCTTTTCTC ATCCATGCTC ACATAGGCC CTTTCTCTTT GAGCTGAGC CTCCATTTTA 1080  
CCTGGCTGAG ACTGTCAAGC TCTTGAGGGA GTGGGCTTCA CTTTCTTCTT GCGCTATGCA 1140  
GTCCAAAGCTT CACTTCTTAC CGAGCAAGGT TGAATCATCT GCTTCAATGT CTCTGAGGCT 1200  
TTCTCTTCTC CTTGAAAGCT AGCTGGGCTG GTCTTGTCTC CAGCTTGTCT CCGCTCTCTC 1260  
GGATGTGGCT TTTCAGGGCC CTGTCTTTCC TCCGCAACA GTTGTCTTGG GTCCATGCG 1320  
AAGCTATCA GGGGCTTTTA CCTTTATC CAAGCATGT AGCACTGCG TGCATTTTAT 1380  
CTCCATCTGG TCACTCTGAG TCCAAAGCT TCAATCCCA TGTTTTCTG GCGCATGAG 1440  
GCTCAAGCTC CATTCTCTG GATGTGGCT GAGCTGTGTT AACCACTTAC ACCCGAGTG 1500  
CTCTTTCTAT GGTGTGGCT GCTATCTG CTATGCAAA CTAGGAAATT AGGCTGATC 1560  
TGGGAGTGGC TTGTGCTGGA GAGTGTCTTT GCTGTCTCTC AGACTCAATC ACTGTGTTCC 1620  
CTCCCGGCTT CTCTTATCTC CATGGCTTTT TCGAGCTCTC CAGGTAACTT TGGGCTTCA 1680

GTGGAATTC CTTTGTGGTT TGTCTCTCTG CTTCCTACT TTGTATTAAG AAGCATTTCA 1740  
CAAGAGAGA GTGGCATCCC TGTCTGTGCT GTCCAGAGCC AGATTTTCTT GAGGGGGCTT 1800  
5 GACCTTACC CTTCAGCTCA GCGCTGTACA CTTGAGCTTG TAAATGATG GGGTTTCTG 1860  
ACTGTATGCC CTGACAGCAG TAAAGCCAA AGGACTCTTG GGGCTCACT GTGAGAGCA 1920  
GGTTTACTTA CTCTGCCAAG TGAAGACAA CTCTTAGGCT GTATCCATA ATTTCAGAT 1980  
10 GAGAAACATT AACATTAATA ATTGTATGTA AACTTAACTT CATGAGACT AAAAAAATA 2040  
AAAAACTTGG GGGGGGGCCC GTAAACCAAT GGGGCGTTTG GGGGGGAGTT TTAATAAT 2098

(2) INFORMATION FOR SEQ ID NO: 149:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1847 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 149:

TGAGCCAG CGTCCGAACT GAGCGCGCG CGAGAGCGCG TTGAGTCTG GTCTTTCGT 60  
30 CCGGCGCGCG GACCAAGCG TCGCGCGCGCG GCGGAGAA GATGTGCTT AGCGGCTCG 120  
GCGCGCGCAC GCGCGCGCAC GATGAGCGC AGCGCGAGCG CGGCGGTGCG CGAGCAAGCT 180  
35 GCGCGCGCTG GCGCGCGCGCG GCGCATGCG CCGCGCGCGCG GGTGAGCTT GATCAAGATA 240  
ATGTTTCA TCAACCGCTT GAGAGCTTG AGGTGTACA TCAAGAGTGC GCTTCCCTG 300  
GTGCTTCTTA TCAATAGGCTT AAGCGCATG GCGATAGCTT TCTGAGCGCT GGGCTACTTC 360  
40 TTCAAAATCA AGGAGATTAA ATCCCGAGAA ATGGAGAGG ATTGGAAATAC TTTTCTGCTA 420  
CGGTTCAATG ATTGTGACTT GTGTGTATCA GAGATGAAA CCTCAAGCA TCTCACAAC 480  
45 GACACCAA CTCGGAAG TACATGAGC AGCGCGAGCG CCGAGGCTTC CAGCCAGTCC 540  
CGCCAGCGCC TGGAGGACTC GGGCGCGGTC ATATCTCTAG TCTCAATCAC CCTAAGCGTG 600  
GACCCACTGA AACCTTGG AGGTATTTCC CGCAACATCA CCAATCTGTA CTCACCAATC 660  
50 TTAGGCAATC AGATTGACTT TTCAAGCGAG GAGCGCAGG AGGAGATTAA CATCACTTTC 720  
ACCTTCTGTA CAGGTGAGG CTCAGATGAC TGGCGCTTCC AGGTCACTG TGAAGAGTG 780  
GTATTTACAG CCGCATGAC CCGCGCGCG AGCGCTGGG TGTTCGCTT CACTGTACAG 840  
55 CAGCGGACT GTTGTCTGTA CAGGTAGAG AGCGCAAGC TCTGTATACA GATCTTGTGA 900  
ACTGCGAGAG ATGCCAAGC AAATATGCGC CAAGATTACA ATCCTTTCTG GTGTATAG 960  
60 GGGCGCTTGG GAAGATCTTA TCAATCTTAA ATGCCAAGC TTACAGTAT TGTTCAGAT 1020

5 GATGACGCTT CATTAATTA TTTCATCTC ATGCACACCA GTTACTTCTT CTATTGTATG 1080  
GTATATACAA TGTTTTGGCTA TGTCTTTATC AAGGGCAGAC CTAGCAAAAT GTCTCAGAGC 1140  
AATCTCGAAT TTGTCCCGA GAGCTGGCT TTGGCTGAG AGCTTAATCCA CAGCTCCCTG 1200  
TTTTTTGAGA GAGACTGAGA GAGCAATAT CTATGCTGCT TGAACTCCAG CTGGGCTCTG 1260  
ATGCTCTCTG AATACATAT CTTCGATCT TGGGTATATC CAGCCAAAGA CATTTCAAGT 1320  
GCTGTAACT GATTTATTA TATTTATTA AATCTAATCA GAAATGGTC CATTAATCCA 1380  
CGTCTTTTC CTGGGTACA CCGAGGCC TTCAACCCA CTTTGGACTT GAGGACTTAC 1440  
CTGATGGGAC GTTTCAGCT GTCTCTAGAG AAGGATCTCT GATCTAGCT GTTCAGAG 1500  
ATGTTTTCAG CAGGTGACA GAGCATTC GTCTGTATG GGTGTGAGT TTGTTTGTGT 1560  
TCTTTTTC ACGCAATATG TAGAGACAT TTGAACAGT CTGCACCTTT GATACGGTAT 1620  
TGATTTTCCA AAGCACCA TCAATTTCT GATTTTATG TGTCTGTGCT TTAAATATCA 1680  
TAGTAAACAC AATTAATCT TTTTCTCAT TTTCCTTCCA GGAACATAC CTTAAGTTTT 1740  
TTTTTTTGTG TTTTGTGTTT TTTGTCTTTT GTTTCTCTTT ATGAGAGAA AATTAATATG 1800  
TACATTTTTA ATACTACCA AATATGACA AAAAAATGCG AGGGGG 1847

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1569 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 150:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 150:

540 GAGCTGTGAG AGAGAGGCC TCTCTCTGA GGTGTGTGC TGTGTGTGAG TGACCTGGC 60  
GGATTAGGCC AACTCGGATC CGCGGTCTGT GAGTCTGGA CCGATCAGA AAGCTGTAGC 120  
CAGCTCTTT CAGGAGGAG TAAATCTCT TTGTGGTTTG GAGCTCTTC AGTTCCTGCT 180  
AGAGGAAGCT CTTTCTGGG CTGTGTAGCC CTGTGATAC ATCGACAGCA GTGATGAGAT 240  
GGATGCGCAG GAGGAGACA TCGATGAGAG AACTGTCTCC AGAAAAAGA AAGCAGAGAG 300  
ACACAAAGA GAACTGAGC GGGCTGAGC AGAAGATAT CCAATGATA TTGTGCTATT 360  
GCTGGCTCTC TATATCTCTC CTAGGAGAT TGTGAATTTT TCCCTGATTT GTTAGAATGC 420  
CTGGACTGTC ACTTGCAGTG CTGCTTTTG GACCAAGTGT TACCGAAGCA CTACAGCGTG 480  
GATGCTGCC TGCCTTTGCG TTGCGACCA GATCAATGCG AAGAGTGGCG CTGTCTCCCG 540

600 GCTTGTGTCA TCCGATCTCT GTACATATG TATGAGCCAT TTCTGTCTCG AATCTCCAG 600  
AATCAGCCA TTCCAGAG CAGCCCGAG ACATTAAGA ATTCCAAATG CTTACTTTTC 660  
5 TGTGTGAGAA AGATTTTGG GAGCAGAG GAGCAATGT GGAATTTCAA CTTCAGTTC 720  
AAAAAGAT CCGTAGATT AAGAGCAG TGTACAGAG GATTCAGGCT TCCCTTTGAG 780  
TAGAGAGAT TTCTAGCAA TCCAGCAG GACTCTGCT TACTCAGGT CACACGCTC 840  
AATTTCTCT TATTTCCAT TGTCTAGGA ATGATATTTA CTCTGTTTAC TATCAATGT 900  
AGCAGGACA TCCCGCATCA TCGATGAGA CTGCTGTCTC AAGATTTCCC TGTCCATGAT 960  
10 GGTTCGAAAC TCCGAGTGA ACGGCTGTG CAGTCAATCC TGAACCCAT GACAGGCTT 1020  
CGGCTCTTTG ACTGTGTGCA TCTCTAGTAC CCAATCTGCC TGAAGCGTGA GTTACTGCTT 1080  
CCCATCCCTT GGGGCGAGCC TCGATGTGAG TCCATTAATA ATCAATTTCC AATTTGAGCA 1140  
20 GGTGTGCTGG ATTTATATC TCGTTAATGA TGTACATGCT CTTCAGGTTT TAGGGCTCTT 1200  
GTTAGGGGAG GAGCAATGT TGAATCAGA GGGAAACAA CTACTATGAT TTATTAACAT 1260  
25 ATTTTAATGT AAAATTTTGC ATTTAAAGG AATGCGCTG TTTTCTGTGT TAAACGCCA 1320  
TTTGTGCTGA TTGATTTGT TCTTTATCT TTATTCGAG TGAATTTGT TGTCTGCT 1380  
GTAGGAGAA ATTAAGCTCT TTGATCTCC AAACAGGAA GTTTGAGCAT TCCCTTATG 1440  
30 ATCAGAGGAA CTTTAGAGC CTGAATTTGT TGTTCGAGT TTAGTGGCC CTCAAATTA 1500  
AGTAAATAT TTCCCTTCTC CTTTAACTT TCTTCAGAA TAAAGCAGGT GACAGGTTT 1560  
35 CAGAACTTT

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1540 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 151:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 151:

60 CCAAGCTTC GGAAGGATT GACAGTTAA CCAATCTTT AGCCCCCATG GCTGTGGCC 60  
AGATTATGAC ATTTGCTGCC CAGTCAATCG GCTGTGGCTT TATTTGCGGA TGGACTTTG 120  
TATCAATGT CTTGAGTAC GTCTGTCTCT GGAAGTTTGA CAGAAACCC CCGCTGTAG 180  
CTGTGAAGC TGTCTTTAAA GAGAGGAAA CTGAATTTGA AAGCTGAT TTACACAAAG 240  
ATACTGAGCC AAAAGCCCTG GAGGAGCTC ATCTAATGGG TGTGAAGAC TGTACATCC 300  
60 ATGAGCTTGA ACATGAGCAA GAGCTACTT GTCCCTCCA GATGGCTGAG CCTTTGCTGA 360

420 CCGTCGAGA TGGATGGGTC TCGTACTACA ACCAGCCCTG GTTCTGGCT GGCATGGGTC  
5 TTGCTTTGCT TTATATGACT GTGCTGGGCT TTGACTGCAAT CACACAGGG TAGGCTTACA  
CTCAGGGACT GAGTGGGTC CATCTGCAAT ATTATGATG GAGCATGAG TATTAATGGA  
ADATGGGAA CTGTACTTTT TACTTGCGTA CTGCGAATAT GTGCTTTGGT TCGGAGGTC  
10 TGAATGAGG ATTGGCAGAG CTTTCTGTAT TGAATCTGTS TGTGATCTCT GTATTCATGC  
CTGGAGGCC CCGGACTG TGCTTTCTC CTTTGTGAGA TATCGGATCA AGCTTCATTC  
15 AAGGAGGTC AATTACCACT ACCAGATAC CTGAATATAC AACTGAAATA TACATCTCTA  
ATGGGCTTAA TTCTGCTAAT ATTGTCCGG AGACAGTCC TGAATCTGTS CCGATATCT  
CTGTGAGTCT GCTGTATTGA GGGGCAATG CTCTAGAAAT GGGTCTTTGG TCCCTTTGAT  
20 TAACTGTGAC ACAGTTGCTG CAGGAATG TAAITGATC TGAAGAGGC ATTATTAATG  
GTGTACAGAA CTGCAATAC TATCTGTG ATCTTCTGCA TTTCATCATG GTCAATCTGG  
CTCCAAATCC TGAAGCTTTT GGGTGTGCTG TATGTATTC AGTCTCTTTT GTGCAATGG  
25 GGCATATAT GTATTTGCGA TTGGCCAAA ATACTCTGG AAACAGCTC TTGTCTTGG  
GTCTGATGC AAAGAGATT AGGAGGAA ATCAGGAAA TACATCTGTT GTTTGAGACA  
30 GTTTACTGTT TCGTATCTG TTACTAGATT ATATAGACA CATGTCTTA TTGTGACTG  
CAGATCTCA ATAAATGCT GGGTGTGTTG CTCTGTTTTT ACCAGCTG TCGCTTGGA  
35 ACTAAGGCT GTTTAGGAAA CCGTAGTGC CAGAAATTA CTGGAATAT TTCCCTTAG  
TTGAGGCGCA TGGRAAAAA ATTGGGAAA GGAATAACT AGTTTAAAT ACCGAGACT  
ATATGGATA ACATGTAAT CCGATTTTC TCAATGATG ATCAATCTT ACCTAAMAGA  
40 GTGGTATGTC ACCTGAATC AGTTATCAT TTGACAGATC

(2) INFORMATION FOR SEQ ID NO: 152:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1719 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 152:

55 TACTATGAG GTCAATGGA AATAGACA CAAATTTACT GGGCTAGGA TTTCAAATAT 60  
TAGATGGC ATGATAGC TTGGTTCAG AACTGTGAT GATGTGGCT GCTGGGCGAG 120  
TAGAGCAAT GTCTAGATC ACCTTTCTG CTGTGATGCT ACTTGTTCG CCAACTCTG 180

ATCTGATCA ACAGGCTGTC GTTCAGGAA TGAATACAGG AATTGAGGA TTATCGAAT 240  
GTCTGGAGC GGGCTCTAT GATTCATATT TCAATATAT CAAATGGA CTTAAGAAC 300  
5 TGGCAATAC AGGACAGAC TTGGAGAA ACAGAGCC TCAGACGAC TTTCAGAGA  
ATTGCAAT CCGTGGGCT CCGTCTCTAT TTGAGGCTG TTCACTACTG CTGGCTCTGC 420  
TTCTGGCTT GTTATTTGG GACATACCA ATTATAGCTT AAGTTCGAG AGTTGGGAA 480  
AGCACTGTGG CAGTCAGAC CATCTGATA ATACAGAGC GCGAGAGAG GCGAAGAAC 540  
CTTTACTCCA GGCACAAAT GTGTAGCAC TGAATCAGG AAGATTTTTC TATCAGACCC 600  
15 CAGCTCTTAG TTTCAGCTC TAGTCTCGA TGTATTTCC AATTCCATCC ACAATGACT 660  
TTAGATCTGT CTTAGAAAT GTATCTGCAAT GAACTCCGTG GAACTAAGG GAATGCGAA 720  
CTTAGACCA GACAGTTTC CAAAGATGTT AATTTCTCT TTGAAAAAC TTTTGTTAT 780  
20 TAGCAACCA TTCTGCGAC TAGCTATTT GTTTATATAT ACATCTCTTA ATTAAAACT 840  
AATATGAA CTCTTAGAT ATTAGCAAT GTCTCTGCTA CCAATTTCTT AAGGTGTGCA 900  
25 GCTTTACTG TATCTGACT CAGTGAGACA CAGTAGTAG TATGCTGTG GACCTATTTG 960  
TTTAACTT GTAAATTTT GAGTCAGAT TTATATGTT AAATCTTGG GTCAATTAAT 1020  
TCAAAGCTT AATCGAGTG CACTAAGCA AAGAAATGTT AATGAAATG TTTCATTTTA 1080  
30 AAAAAAAA CTCTTAGAA AACTGACTA AATCTGATC ATGTTTTTGG CTGTGTTGCA 1140  
GTACTTTTAA ACATATCTA CTACTGTTT TGAATGAGA AAGTATCAG CATTTAGCT 1200  
35 TTAGTTGGG GTATTTAGG CCGTAAATCT AAATCTGGC TCAAAATTA TCCCAAGCTA 1260  
CTCTATAC CACTATCTT TTATGTTTG CAAATGATA AGCACTGCA CACTTGATA 1320  
40 CATAATCTAA AATATATATA GTAAAGCTGG TAGCTTGA AATCTAGTG TGAATCTAT 1380  
TATGTAGATA AATATATATA GTGGCTTTC AGCACTGCA CAGTAACACT TTATTTTACG 1440  
ACCTATGTT TGTCTTAAT TTTCAGGAC CTAGAGAGA GCTTTATACA ATTACCAATG 1500  
45 TGAATTTCTC TAAAGTGTAT ATTTTGTGT CAGTTATAT TATTTAATA AGTGTACTT 1560  
TGTAAATAT GTATATAGG AACTATAG TTTCACATGT TTTCATCTTG TGTGTGTTA 1620  
50 TTGCTTAATG CTTTATTAAC TTGGAACT CACTATGTT AATATAGCTC TTAAGAGAA 1680  
TGTAAATAT TTGTATATA AGTAAATAT TTATATGAT 1719

(2) INFORMATION FOR SEQ ID NO: 153:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 863 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 153:

5 GGCACGAGGG AAGCGGGGAC GATGTCCGCA TCACACCCGA CTTTGAGTTT TGGAGTGTCT 60  
TGCCTTAGAG CAGGCGAAGC AGCTCTCTATT CAAAGGAAGT AGAGGCTCTT CCGTCAGTGG 120  
TAGCGAGACA GCGAGAGCG GTTTTCTGGG AACTGTGGGA TGTGCCCTTG GGGGCCCGAG 180  
AAACACGAAG CAGAGTCTTC CAGACAGTA ACTACAGCTT GGTGCTCTCT CTGCAATGCC 240  
TGTGCTGTTC CTATGACTTC TTGTGCATTT CTTTCTGGA ACTGTCCGA AGAGCTGCTG 300  
TCATCCAGCT TGTGCTCTTC ATCATCCAGG ATATTCAGT CCTCTTCAGC ATCATCATCA 360  
TTTTCCTCAT GTTCTTCAGC AACTTCTGCT TCCAGGCTGG CCTGTGTCAC CTCTATTTCC 420  
ATAAGATTCA AGCGACATTC ATCTGACAG CTGTGTACTT TCCCTCAGC ATCTGCTTTC 480  
ATGTCTGGGT CATGAACTTA CCGTCGAAA ACTCCAGAG CTTCTATGG ACGAGTGGAC 540  
TTCAATAGCT GTTTATATTC CAGAGACTAG CAGCAGTGTT GTACTCTTAC TTCTATTAAC 600  
GGACAGCCTT AAGACTAGCG GATCTCTACT TCTACAGGA CTCTTTGTGG CTGCGAAGG 660  
AGTTCAATCA AGTTGGAAGG TGACCTCTTG TCACACTGAT GATACTTTT CTTTCTGGA 720  
TAGAGGGCCA CATTGTGTCG TTTGCAAGGG AGATGTGGCG CATTGTCAGT GCGCAAAACA 780  
GGTGGGATTT TCCAGGGAAA GGGTTCAAAA TTAGGCTGTT TGTTCAGGCC ATTTCAGG 840  
AAGCGAAGG GTTTCCTGTC CCT 863

(2) INFORMATION FOR SEQ ID NO: 154:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1101 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 154:

50 AACAGCAAAA AAGATGATTT TCTTCTGAAA TTGTGGAACA TGAGTAATCA AGTTTATTT 60  
TTGTACTAG CTCTGTGAGG AAGATCCGAG TTCCAAAGCG CCGCATCTCT TCCCTCTGAG 120  
CCAGAGCTCG CGGTGGATTC AAGTCCACT GAACATCAG ACGAATGAG ACGAATGTA 180  
GAATGATGAT AATGACCCCA AATTAAGGTT TTCTTGGGCG AGGATGTGCT GGAATGAGAA 240  
AGGTGATGAG ACGAGGAG AGCAGATCG CAGCCAGCAC AAGATGAGT GTGTGTTATT 300  
ACGAGAGGCC AGCATTTGAG CTTAAGGTCC TTCTACTTAC CTGATTTGCG CATTGAGGT 360

CGAAACCTT CTACTGCC ATAGCCAGG AAAAGTGAAA AGAGAACACA GTTCTTTTAA 420  
GAACTCCGAG CAGGCTCTGA GCGTTATGTT ATGTAGCTGA GTCCAGCAGG TACATGATGC 480  
TGTCTGCTTT CAAAAGACT TTCTCTCTCT AGCTGACTGA CTCTTCTCTT AGTTCAGAGA 540  
ACAGCTGAGA CAGAGCTCTG CTGAGTAGCT CTGTGATGAC AAAGCTTTGG TTTAAGCTGAG 600  
GTGATCTCCA GGTGTGTCAGG TTATATTAGTC CCGAAGGCA ACGAATATAT TAGATTAATA 660  
ATCCAACTTT AATAGTATAC ATTTAAGGA AAAAAGACAA AAGGCTCTGA AGTATTGAGGC 720  
CAAGCTTCTT GAGTATTCOA GCTGCAATTC CCGAAGGGA ATCCAGACA AGTCCCTGCC 780  
TGTATTTTGT TCTTGAGAGG GGTCAATCTA GAAGCTAGAT CTTATCAGGA TGAAGAGCAG 840  
CAGCCACAGG CTGTCTCTGA TCAGCAGCAA CGATTTTAAA GAAAAGAGA AGAGTTTCTT 900  
AGTGAATGTA TTGTATTGTA AGATAGTCAG TGAATACAC TGAACAGATG CTATCATATC 960  
ACTATGTGTC CTTTATTAGA TAAAGTTTAC ATATCATCAT TCTTTTGGGG AAAATTTGTTA 1020  
TTCAAGTATA AAACAGAG AGTATTAATA AAAATTAATA GAAGCTTAAA AAAAAAAGAC 1080  
CTGTGCCGA ATTCCTCTCA G 1101

(2) INFORMATION FOR SEQ ID NO: 155:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2031 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 155:

40 CAATTAAGCC GTTTCAGGCC TAGGTTTGTT GCGAAGCCCG NCGCCTAAGG TTTTAAATGG 60  
GCAGACCAA GCGCCTGAAA GGAAGGGAAA GCGGAGGGTA GCGGAGGGTT AGCAGGTGAG 120  
TTCTTAAGGC TGGAAAGGTT AGCAGAGCC TGGTGCATG CCGTGTGATC AAGACAAAGC 180  
CAGGATCTTC CTGGTGGCTT ACCAGCTTGG GTTTTATCAA AAGCAGGCTG GAGTCTTATT 240  
TTTGTACATG AGATACATCA CACTTACTG TGGGCCAGTA TTGTGAGCTG AGTCTGAGTT 300  
GTTTTACTG ATCCCTTCCC TGCCACACAC AATTTGTGTA CATATCTTTC AAGATGATAC 360  
CAGCCCTTTC CCGAGCTCCC AACCAAGAC TGGTTCATAG CCGTGTGTTAT ATGTCAATTT 420  
TAGCGTTTTT ATATATGACC TTGTATTTCT GTTGTGTTTA TTTTAAGACA GTGTATGAC 480  
CTTCATTTAA ATACATCTCT GTGCATACAG ATACGCATAT ATGTGTGTGC GTATGCATAT 540  
ATGTGTGATC TGTAGTTTCC AAGATTCAG CTGAACGAGA TGAATGCTCG CAGCCGAGAA 600  
GACAGCCCTG ATCCCTCTCA ATAGTGTGTT CCAAGATAT TATGAGTCTT TCCTTATTAA 660

5 TATTTTCAT TCAGAGACT GAGCAGAGC TGATATGTTT TGTCTTTTCT TTGGAGCTA 720  
AGTGAAGGTC TTGGATGAC TTGCTGTGTT CTTGAGGTC GACTTTGGGG CACTCTCTCC 780  
AGTATTAAGC CCGCTTTTTC CTGTTGGTGA CTCTCTCTCT GCTCTCTCTCT GTGTGTGATA 840  
GTGACTCTTG CATGAGCTTC ATGTCTGTTT TGTGGCATTT GGGGATTAAT GCTGAAGCAG 900  
10 AGCATTTTGA GTTTTGTGGA GGGCTCTTTC CAAATGATAG ATCATCTCTG TTGAGCTGGT 960  
ATGTCTCTTT GCTTGTCTCT TTCTCTCTCT TTCTCTCTGA AGAGGAAAGG ACTCTGTCTA 1020  
GGCCAGGCTT GATGAGATG AGCTGAGCTT GCTCTATGCG GTTCTTATAG CAGAGAGAG 1080  
15 AGTATGTGAT TTATCTAAT TCCTTAACAA ACATTTATTC AGGCAAGCTT CTTTGCATAT 1140  
CCAGAACTTG AGGCAAGCTT GGGTATGAC TTGCTGATGA ATATATAGCT GCTAGAGGGT 1200  
20 AAATGAGGTC ATCAGATTT CTGTTCAGCG GGCAGGAATA GCTCTGATAT TGTAGCACT 1260  
TTTTTTTATA GCATTAATTT TTGACTTGT TTCTCTCTGAA GTGCAAGAGG CTTAGACTTT 1320  
TTCCAAATGT AGACTAGAT CTGAGAGT CCAGCCACTT TATATGTTAT TGGCCAGCCC TGCTTTTCTC 1380  
25 GAGGAGAAC TTTTGGAAAC CAATGATCT TATATGTTAT TGGCCAGCCC TGCTTTTCTC 1440  
GGTTGAAA TTGACTAGT GATATGTTT TTAGAGAGAG TGTCTTGGAC CATGGGTTAA 1500  
30 CAGGAAAGGC TACTTAATTT CACTATCTG CAGCAGAGC AGCCAGCAG CATTTACTAG 1560  
CAGCAGGAAA ATGATGTAT TTGAGTTCTT GTGTGTGCA AACTGAGGCA CCAATTTCTT 1620  
TGAAACATG CCAGCTCAG GCTGGGGGCG GTGGCTGACA CTTGTATATC CCAGCACTTT 1680  
35 GCGAGGCGCA GCGGGGGGGA TGACCGGAGT GGGGAGTTT GAGACAGGC TGGAGCAACA 1740  
TGGAGGAAC CCATCTCTTA CTTAAATATA CAATTTAGC GGGGCTTGT GGCATGCGCC 1800  
40 TATATATCCA GCTACTTGGG AGGGTTGAG CAGGGAATTT GCTTGAACCC GGGAGGGCG 1860  
AGGTTTGGCG TTGAGTTGAG GATGTGTCCA TTGCACTTCC GGGCTTTGG GCAACAGAG 1920  
45 CAAAATTTCC GTTCTCAAAW HTGTGCGAAT TGGATATCAA GCTTATGCAAT ACCGTGAGC 1980  
TGGAGGGGCG GCGCGGTACC CAATTCGCC TATAGATATC GTATTAATAT C 2031

(2) INFORMATION FOR SEQ ID NO: 156:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1981 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(41) SEQUENCE DESCRIPTION: SEQ ID NO: 156:

5 CTTGCACTCT GAGCCCTTCA CCGCTTCGAG TTCCCGCCAG GTTGGCTTCC TTGATTTCTT 60  
TTTCTTGCTA TCAGATTTTG ATTGAGAA CAACCCCTTC TTGTGACAG TCATATATGA 120  
20 GCTCACTCTG GAGAGGAGC TGGGGCAGAG CTGACCTTGA TGTTCATTTT TACCCCGCAA 180  
GACCGGAAA ACTCTGTCTG GACAGAGAT GGAACCTCAG CCCAGATCTT GAGCAGACTT 240  
10 GTGTGCTCTG CAGCTGAGC CTTGCCATG TTAGAGAGC AGCTCATAGA TCCCGCGGGA 300  
CCTGGGGA CA TGAAGAGT GTTCCGGCGG CCGTTGACA TTATAGAGT GCTGATTTCC 360  
CTGTTCCTTC GCAATATCC GGGGCAAGC AGCTTTCTG ACTGCGAGT TGCCTCTCTC 420  
15 TCCCGGGGCG TGTCTAGCA CCGCGGGCCC TCATCCCTGA TCCCTCTCTT GGTATATAT 480  
CTTCTCAGC TCTATCTGAC GAGCTCAGG GAGGCGTTTG GGAATCTGAG CTTTTCCTTC 540  
TATGACAGC ATGATGAGGA GGTGATTTGT GTCTCTGGA AGCCCAAGCAG CTTTCAAGCG 600  
20 CAGCCCTTCA AGGCTTCAG CACAAAGGGG GCAATGCTGA TGTCTCGAGG TGGGAGCTA 660  
GTATGTCTCC CCAATGTGTA AGCAATCTTC GAGGACTTTG CTGTCTGAG TGAAGGCTTC 720  
25 GTGCACTCTG TGAAGGCGCG AATGAGAGG TGGACTTTGT GATCCAGCT CTGAGCAGAG 780  
CTGTAGAGCG ACAGCAGGAC ATTGAGCTTC TTAGAGAGGA TGTGATGAGG ATGAGCTTCA 840  
30 CCGCTCTCTG ACATGATTC TCAATGAGG GCGTCTGAG TGAATATCT GAATCATCTC 900  
CAACAAAGC CAGCCCGAAC TTCTCTCTCG ATCTCTGAG ATTGGGAGG GGGCATGCTG 960  
GCGCATGAG TCTCTCTGCG CTACCAATTC CAGAGAGGA GTGGGAGCCA GCTCAAGAAA 1020  
35 GGAATGAGC CCAAGAGATC CATCAACTTA TTAGCGCTTG GCTTGAAGCT CCGTGGATTT 1080  
TCCCACTCTT TTCTTATGCT TCTTTCAGAA ACAGAGAGG GATGTGTCTC CTGGAGAGG 1140  
40 CTCTGTCTCC TTCTCTCTG CAGGAGCTGT GCTTAGACTT AGCATGCGCT TCACTGCACT 1200  
GTGAGGCTTT TGAATGAGAC CAGCGGAAAA TGTGGGCTTT CTGATTCACA TCACCGACAC 1260  
TGAAGAGTGG AAAGGGGCTA TATGTGTATG ATTAGAGCAC ATTGAGAGG CACAATGCC 1320  
45 TCTGTGTGTT ATGCCACTTC CAGAGGTGGA GACATGGA AAAGAGGAG GACAGAGAG 1380  
GATTTGGTAG GTGAAGGGTT CAGGGGAGTG GTAGTCAGCC ATCTCTGAG AGGTGCAAAA 1440  
1500 AGCACTGGGG GCTTACCGCTT AGCTGATCTT GCGTGTGCTG TTGCGCGCTT CATGTCAAAA  
50 ACTGCCACTA CTATGTACTT GCAATGGGGT TGCAGAGATG GGGGAGACTC AAGTCTTACT 1560  
CCCGAGAGC TCCCAAGGCG CAGAGAGGAG AATGTGCTCT CTTTTCAGTC TGGTCTACAC 1620  
55 CCAATTTCTG GTAGCTCTG TGTCTCTCTT AATCTGTGCT GTTTTTCAG ACTGAGCTCA 1680  
AATAGTCCC CTGCTTATAG CAAATCCCTG CCGCCAGGCT GAGGTGATCT TTCTCTCTCC 1740  
TGAATATTA GAGCAATTAC TGTCTCTTCA GTTGTCTTGG CAGGCAACA CAGTGGGATA 1800

AAATCTATTG TTTTGACTC TGATTAAATA TTAATTGCA GCTGGGGTG GTGGCTCATG 1860  
CTTGTAAATCC CAGACTTAG GAGTNAGAG GAATCACTG ASCYCAGAG TCTGAGACCA 1920  
ATCTGGGCAA NAGAGAGACC CCATCTCTT TAAATAAATA GTTAATTGCG TTAATAAATA 1980  
A 1981

## (2) INFORMATION FOR SEQ ID NO: 157:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 157:

GAATTCGCA CAGCGCGGC CATGGGCTC GTGGTTTCGG TCGTGGGTGT ACTGCTGGGC 60  
GGCTCTTCG GGTCTGTGGG GTTGGCCAG CTCCTGAGG AGATCTCGGC TCCATTTTCG 120  
GAGCGGATGA ATGCTCTGTT GTGCAAGTT GCTGAGGTGT TCCCGCTGAA GTTATTTGCG 180  
TACAGCCGAG ATCCCTCGAA CTACAAATA GCTGTGGCT TTCTGGAACT GCTGGCTGGG 240  
TTCTGCTCGG TCAATGGGCC ACCGATGCG CAGAGATCA GTAACTGTGT CTTGATTCG 300  
CTCATGATGG GGGCTATCTT CAGCTGSCA GCTCTGAAG AGTCACTAAG CAGCTGTATC 360  
CCAGCCATGG TCTGCTCGGG GTTCTGCTG CTGCTGATG TCGGCGAGCT CTTAGCCGAG 420  
ACTAAGAGGG TGGTCAGACC CACTAGAGAG AAGACTCTAA GTACATTCOA GGAATCTCGG 480  
AAGTAGAGCA TCTGTGCTC TTTATGGCAT GAGCTGTCA CAGCGAGAC ATGGTAGAAC 540  
ACAGAGTCTA TCATCTGTT ACCAGTAAAT TATCCAGGCT CAGCCAGTGT TGAAGAGAGC 600  
ATTTGTCTTA CTTGGCACTG CTTTCTCTTT TTAGCTTTAC TACTCTTTTG TGAAGAGTAC 660  
ATGTTATGCA TATTAACTT CTTCAATGCA TATCAAAATA CAAATTAAGC AGAAAGGAAA 720  
TTTAATTCOA CCAAAATCT GATGCCCAA ATACCCACTT TTAAATGCTT GTGTATAGTA 780  
TACTCTGAA CTTTATTCTG TGCCTTTAA CAGATATATA TTTTITTTTA ATCAAAATAA 840  
AACCATATAT CATTATTAAT TCTCTCTTT TAAACCTTA TAAACTATTA TAAATAAATA 900  
AAAAAAAAA CTCGA 915

## (2) INFORMATION FOR SEQ ID NO: 158:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2117 base pairs

- (B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 158:

AGAGCGAGGC GAGGTGGCG CCGTTCGCG CATGAGCTG GCGCGGCGCG TCGTGGCGCT 60  
GCTGCTGCTG GCGCGCTCG TGGTGCAGGC GGTGAGGCGC ATCAGCTCG GACTGGCGCT 120  
GGCGCGGCTC CTCAGCGCT ACHCTACCC GCTCTCTAC TCGTCTTTG CCGATGCTG 180  
CGGCGAGAG CCGAGCTTTA GCGCGGAGGC ACTCGAGAG GATCTGAGC ACAACTCTT 240  
TGGACGCAAT CTTGCAAGA AATCAATCTT AATGCGCTG TTTGTTTCA TAAACAGCC 300  
AAGGCCAGG AATCTCTCA CCGTCTGCTT GCACGCTG ACAGCGAGC GCAAAATTT 360  
GCTCAGCAG ATCATCGAG AGATATTTA CGAGGCTGT CTGACAGTG ACTATGTCA 420  
CCTGTTTGG GCGCAATGC ACTTTCACA TCGTTCMAAC ATCACTTGT ACAAGATCA 480  
GTTACAGTTG TGGATTCAG GCAAGCTGAG TCGCTGTGCG AGGTCCATCT TCAATTTGA 540  
TGAATTCGAT AAGATGCAAT CAGCGCTCAT AGATGCCATC AAGCCTTTCC TCGACTATTA 600  
TGAAGCTG GTGCGGTCT CTTACAGAA AGCAATGTC ATATTTCTCA GCAATGCTCG 660  
ACAGAAAGG ATCAGAGTG TCGCTTTGGA TTTCTGAGG AGTGGAAAGC AGAGGAGAA 720  
CATCAGCTC AAGCAGATG AAGCGGCTT GTCTGTGTG GTTTTCATA ACAAGAGAG 780  
TGGCTCTG CAGAGCACT TAATTGAGC GAGCTCATT GATTATTTG TTTCCCTCT 840  
CGCCCTGGA TACAAACAC TAAAAATGT TATCGAGTG GAATTCGAT CCGAGCTTA 900  
TGAATTCAT GAAGACATG TAGCAGAGT GGTGAGGAG ATGACATTTT TCCCGAAGA 960  
GGAGAGGTT TTCTCAGATA AAGCTGCAA AAGGTGTTT ACCAGTTAG ATTATTTACT 1020  
CGATGATTA CAGTCATGAT TGGAGCGCG AGTCACTGCC TGGAGTTGGA AAGAAACAA 1080  
CACTGATGC TTCCAGACTT CAGCCCGAG CTCCTTTGCC TGGAGAGGA ATCCATGTA 1140  
TGTGCTGTT TGAATGTACA GGAATTTCC CTGGCAATGT TCCACCCCG TGGTGGCTGC 1200  
AGCGACCCA GCGACGAGG GCGAGAGCT GAGGCTGCC GAGACGAC AGAGAGAGG 1260  
AGCGAGTCC CAGCGCACT ATGCGAGGC TCATGATTT TTACAAATTA TGTTTTAAT 1320  
CGAGTGTG CTTTTCGAG GAGGATGAA TAGTTTATAT TGAATATG GTACATTTAT 1380  
TTAAATGAT TTTTAACAT ATGAGAGCT GCTCAGATTC TAGTTGTG GCTTTGTG 1440  
TGTGTTTTT TTTAGTTCT CATCATTTT ACATGAGTG TATGTATCT TTACTGAAA 1500  
TAGCGCAG CAGCATGCA TGGCATTTT TCCACAGAG GCAATGCCG GGAATGTGC 1560  
TGGACATCA GCGAGCTG TCCAGAGTG GTCCAGCGG ATGCTGCCAG GGCATTTAA 1620



5 GTGTTAGGT GAGCAGAG TAGTAGAG GAGCGCTTCA GGCACACAG ATAGCGCTGA 1680  
AAGAGCTCT CGAAGGTTT TCACTTAGC AAGATGGGA GCTGTGGAG TGAATTTGCG 1740  
CAGACTGTCA ACATTTGTGA GAGCGTCTT TTGGAAGGA AAGTAATTTCC AACTTTGAC 1800  
TTGCGATCA CTCGTTTTG CAGAGTGGG CCGTTCACTG TCCATTTGCA ATAGCCGCA 1860  
CGTGCCTCT GCTGATCTT AATATGTG AATTTGGCA TATTAAATCT TCCCTATTGA 1920  
TACTATATT TGTTAGTCT ATTCAGATC CCGAAGCT CCGATTAAGC TTAGCTGCCC 1980  
CTTCTGAGA TCGTAGAGC GGTGTCTTC TTATTAATG CAAATGGCTA CCGTTTACA 2040  
15 ATAAATTTT GCAATGCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA GGGGGGGCG 2100  
GTAAAGATT TGNCCC 2117

20

(2) INFORMATION FOR SEQ ID NO: 159:

- 25 (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2195 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 159:

50 TGTTCCTTAA TCCCTTTTCT AAAAAAGGGG GAAAAATCCGG ATGCAATTTTA GCGATGCTC 60  
TGTGTGACG TGTGTTTAT TGCACCTA ATCTCGATT ATAGGCTTTT CATTCTCGG 120  
CAAGGCTTT ATTTTGGCAG TTAGCGAAA TGTGTTTTC AGAAGTTAG TTATTTCTC 180  
CTCTTGTTT CTTTCTTTC CTGCTTTT CTGCTGAC CCGAAGCTT ATGTCGAAA 240  
40 CAGTACTGGA CAGAGCTTT TGTTCCTGA CCGTGAATA TCAAGTCTG CTAAATATGA 300  
CAGAGGTGC AGTTTGTGG TTATGTGCT GATTTGCTT ATCAATCAT ATTAGCAGA 360  
AAAAAAGA CTTGTTCTG TTGTACTGA GTCTTAAGA AAGTGGCCC ATAGTTTAT 420  
45 GCGAATTC CAAAGCTTT AGTACACT GTATTTGAAA ATGGGGGACC CAACTCCCG 480  
GAGGAACA GCTCTGAACA GACTAGTGC TCACTTAGA AAAAAAAAAA ATCTCTAGC 540  
50 TGACCACTTT GACTTCAGA TGTATGTC CTTTGTATC AAGGAGAGA AGAAGATGC 600  
AGCACTTTT GAAGTCTG AGTATPAC AGTCATGCA ATTAATATG AAGAAATAT 660  
55 CCGAAGGT GTCCGATT CAGCTATC CTTGGAAAT TCCCTAGAGC TTATTACGAA 720  
GAGTGTGTA CAGCTCCATG CTCTCGATA TCACTATAG AGAAGGATG TAATGCTG 780  
TACTAGAGAG ATGCAATTA TCTTTGCC TCGGAATAT ATGAAAAA TCCCTCTCT 840  
60 CCTGTTTCT AGGTGGAGG ATCTGATCA GCTTTTAGG CCGTTTACG CAAATTTAG 900

5 TTTCATGTC GTGACTATGA AAACAGTTT CTGACTATAC TGACCGGCA GCAAGAGCT 960  
GGATCGTGT TCACATATCC TACCACTCA GTTTTCTCT TCATTGACAG ACAGACTTG 1020  
CAGACTCAA AAACAGAGC TACATCTTC AAGTATGCA GCATCTGCTT CTACTGCCA 1080  
CAGAGAGC TCACCACTG GGGCAGTGG CACCTAGAG GTCACCTCC GTCTTATAT 1140  
10 GGCAGAGTAG AGTACTGACC AGCAAAATG AGAATGATG AGAATGATG AGAATGATG 1200  
TTTCTGTTT TCTTACACT TATCTCTTC AGATTTTAA GAATATGAG TCATGACAG 1260  
AAGCTATGC ATTTTGAAC TTGTATCC TGGATTTT TAAATCATTT TTATCTGCA 1320  
15 ACTTAACAA AATATGATG TGTGATCC ACTGTGCA AGAAGTCT TACACTGAAA 1380  
CCTGCACTG ATCAGTATCT TACTAAAAAT GTGAATGAA AGCACTATG TACACTGAAA 1440  
TGCCTAATG TATCTGAAG CAGAGGTGA TACTCATTTT TATGCTCTC CCAATTTGCG 1500  
20 TGTGTTTTC CTCTTGACA TCTGTATCA GTATTTAGAG GTGAGAGAT GAATGTACA 1560  
GGTATTAATA ACATTTTAA AAGCATTAC TTGTCTATA TCACAGTTGT TCCAGAGCAG 1620  
25 TGTGATAC ATTCATATGA CCGAGACTGG TTTAAAAAA GAATATGAA CCAATGGAAA 1680  
GAATCTTAA ATGAAAAGG CATCTCATG TAGGCATTTT TCCCTCATAT TTACTGCGC 1740  
30 CAGTTTGT TCCGTGACT CATGTATTTT TTTTTCGAG ATCTCTTTC CCAATTTGCT 1800  
ATTGTAGAG TATCTGCTG CCGTGTGAG CAGTTATGA CATTAAGCA GATCTGAGAT 1860  
CTGAAGTAC TATTAAGAG CTATTAACA GAATATGAT CATGACTGCA GAATCATGA 1920  
35 TAGTGAGAG ACTTTCCTT TGTGTTGTT TGTGTTGTT TGTGTTGTT TGTGTTGTT 1980  
CAGAGAGAG ATTTTATTA CAAAGAAAA AATTCAGT ANTGTGAG AATGCTGCT 2040  
40 TTTTACCA TCCTAAGAA AACTTTACA AGGTGTTTT GAGTATGAA AAGGTTTAT 2100  
AAGTGGAT CTTAATTTT AAATTAACC ATGAGTCT AAGTCTTAA AAGCAGAGCT 2160  
45 CATTTGTC ATGAGATA AGAAGAGCT ACTGTAGAG TTTTTCCTT TCTCTCTTT 2220  
AATGAGAA AGCTTTGCT TAGGGTTTC ATACTTTAT TCGATTAAT CTGAATGATC 2280  
CTACTCTTT GCAATTAAC TAGTGTATC CAGTTTCCA TTGTATTTAG CTTCTGTTG 2340  
50 GAATTCAAA AAAAAAGAA AAAAAAGAA GAAGACTTA ATAAATAGG TGAAA 2395

55 (2) INFORMATION FOR SEQ ID NO: 160:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2120 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 160:

5 CCGCGATAC CGCTGAGCT AGTGCATATC ACACCTCTCG CGTCTCGCG CGTCTCGAGCC 60  
TAATGAGAC GCTTGGGAA AGGCACTAAC GGAATTTCCG GTGACCTTC GCTTTTACGGC 120  
TGTGATGTC TTCCGCCAA CCCAGAGAA GCGGAGAGC AGTTTACAC AGCGCCGCTC 180  
GTGTTTACGG CGGCGCCGCG TCGCGCCGCA TGTTTCTCTT TTCTCTGTT TCTCAAGAT 240  
GCTGCTGCTA ACCGCGTCCC CGGCGCGCAC CATCTGTTGC CATTCGCGCC GCGCGAGGCA 300  
TTCCAGATT TTGAGATGG CAAGTTTCAT GACACCGCTG ATCCAGACA ACCGCTCAG 360  
CTGCGTCCC TGTGCGGTTT CCGAGCATTT TCGGATATG CCTTACAGC GATTACGAA 420  
AGGAGATCGG CTAGGAAAGG TTGCGAGCTG GACAGGAGCC ACATACGAG ATAGAGGTA 480  
CACAAATAG TACTCTCTC AGTTTGCTG TGGAGTCAA TATGCTTAT TTCTATGAGA 540  
GGATGAAT AGCTTCAGC TGTGCTATAC AGCGCGACA CAGAGACCG CTTACCAGCG 600  
GATTCAGTG AGTTTCCC AGGAGAACT CCGCGAGAC AAGATGTC GBAACATGTT 660  
CGAGTTTAC CTGAGATCC TCCCTAAGG TCCCAAGCG AAGAGAGAG AGCGATTCG 720  
ACTGCGAAG AGTTTCCGA ACAAATTTG GATTAGGCG AATGCGATC AGAATTCGA 780  
GAAGACCGCA GACTCTTAC TTGAGTTGG TAGTGATTTG GAATGAAAG AGCAATGGA 840  
TTTCTCTCAG TTGATGAGA TCCCTACTT GGAATGATCA GAGCGAGCG AATTTGATG 900  
TTTGTGGGCG CTAGATACT AGCAAAAGC GTTTGACCG ATCACACGA GAGTGAGAA 960  
GCGACTGCGG ASATGAGCG GCATCTTCA CACTGTGAGC AGCAGAGCG ACCGTCTCAT 1020  
CGCGAAGCTG GCAAAAGCT AGGGAAATG GTTTGCGACT GATGCGATCC TGGCCAGGCT 1080  
GATGAGCTGT ACCGCTCAG TGTATTCCTG GGAATTTGTC GTTCAGAGAG TTGCGTCCAA 1140  
ACTCTTCTTT GACAGAGAG ACACTCTGA GTTTGACCTC CTGACATGA GTGAGACTGC 1200  
CAATGAGGCC CCTGAGTNG AGGTATATTC CTTCATTTCA CCGCGAACC TGGCCTTGA 1260  
GCGAACTTAC ATACACGA ATTTCTCCA GCAATGCTTG AGAATGCGGA AGCAAGATA 1320  
CAACTTCCC AACCGAACC CGTTTGTGGA GAGGACATNG GATAGATNG AATATGCTTC 1380  
TGTTCGTAC CTATACCGA GTGGAAGCT TGGAGATGAT ATTGACTTGA TTGCTCGTTG 1440  
TGAGCAGAT GCGTCTATCA CTGAGCGAA CCGGGAATNG TCTTCTATCA ACATCAGAC 1500  
ACTCAATGAG TGGATTTCA GGCATCTTAA TGGGTTTAC TGGCTTGAAG AGCTGACTC 1560  
TGAGGAGGG GCTGTCTATG CAGCGAGCT GAGGAGAAC AGCTAGCAT TGGCGCGCTG 1620  
GACCTGCTGT GCTTTGCTCG CTGATCTGA GTACCTCAG CTGCTTATG TGTCTCGGTA 1680

60

CGAGTGAAG GACTGCTTAC GGCAGCTCAT CTTAGGCAAC CAGAGTTCA AGCTTAATGA 1740  
GTTTCCGAC CAGATCAACC TGAGCTTGA GATGCTCTGG GCGATTTTAC GCTGCTTCAT 1800  
TGAATCTCG ATGAAGCTGG AGAGGGGCA ATACTGATC CTCAGGACC CCAACAGCA 1860  
GCTATCTCT GTCTACAGCC TCCCTGATGG GACTTACG TGTATGAG ATGAGAGGA 1920  
AGAGGAGG GAGAGAGG AAGAGAGGA GAGAGAACT TAAACAGTG ATGTGAGCT 1980  
GGAATTTTTC CTTCACCGA GACTACGAG GCTTTGATG CTTAGTGA TGTGTCTA 2040  
ACTGCTCTC TGACATTTAG CAGATGAAT AATATATTA TCTGTTACT CTTAATAAAA 2100  
AAAAAAAA AAAAAAAAAA 2120

(2) INFORMATION FOR SEQ ID NO: 161:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 900 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 161:

30 GGAAGCTGAA GTCTTTCCAG ACCAGGACA ACCAGGCAAT TCTCTATGAA GCTCGACCA 60  
CCTCGACCT GACTGTATAC TGAGAGAC AGAAGCAAAA GTTCTCACTC AATCGATAG 120  
CGAAGATGG GCGTTTCTTC ATTAGAGAGA ACTTCTTCCA GCGGCGGCC AGGCTCTGC 180  
AAGTCAGAA GTGAGAGAG GTTATCTGA CCGCACTCT GGCATCTCT ACTTCGATGG 240  
GTTTCTGCT TCACAGGAC AATACAGAT TCTTGTATTT ACCGAGCTG GGGAGAGGC 300  
TTGAGTGGC CTTGAGATTC AGCGCAAGG ATGTCTGTG CAGAGATGC TGTCTGAG 360  
GTGCTGTC GCTCTCTGA TCGCTGTGAG TTCTCTCATG AGAATGATTA TTTTATGGA 420  
AATGTGAGG CTGAATATAT GTTTGTGAT CTAGAGGAC AGATCTAGAT GACTTTTGA 480  
GCTATGCTT TCGGATTCG GTATTTGCGA AATGCGAAG ACCTGCGTCA CTTGAGAGC 540  
ACGAGAGCC CTCAGAGGG GBACTCTGAG TTGATGAGA TGAAGCTGCA CAAGGGAATGC 600  
GGGCGCTCC GCGCGGCGA CTTCCAGAGC CTGCGTACT GCATCTGAA GTGCTCTAC 660  
GGGTTTTCG CATGAGAAA TTGCTTCCC AATATGAGG AATCTATGA GCAAAAGAG 720  
AAGTTTGTG ATAGCGCGGG GCGCTTCTG GAGCGCTGCG GTCACTGAT CAGGCGCTCA 780  
GAGAGCTTC AGAGTATCT GAGGTGTGAT ATGCGCTTCA CATTATGAG GAGAGCGCC 840  
TACCGATCG TGAGAGAAA CTTAGAGCT TTGCTGAGG ATCTGCGTGT GTCTCTCAT 900

60

## (2) INFORMATION FOR SEQ ID NO: 162:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 162:

5 GGCACAGAT GAGGGGAC CAGTCTTCT AGGGAGGCT GGGTGGTGGT CCGCTAGTA 60  
15 TGAGCCCTC TTACTGTACT CTCGGGAAT GTTAACTTTT CTATTTTCAG CTTGTGACAC 120  
CTGTCTAGGC AGCTGGCTT CCGATTGGC CCGTGTGGGT CCGACGACG CTGGCTGGCC 180  
20 CCGAGGGCA CCGCTCTCTT CTGTACTTC TTCTCTAAC AGTGACTTGG GCTTGAATCT 240  
GGCAGGAC CTTCCTTTTA GTTCCACAC CAGGAGAGA GTTGTACATG ACCTCCCGCC 300  
CCCTCGACA AGCTGGGAA CAGAGGGAT GTGTGAGAG CAGAGTTCTT CTGGCCCTCT 360  
25 CCGAGGTGTT TTCCACTAGT CACTACTGTC TTCTCTTTGT AGTAAATCA TCAATATCT 420  
TCCCTTGCTT GTGGGAGTG GAGAGGCTGC TGGGTGTAG CTGACCTTCC CCACTAGTT 480  
GGGAGAGAG GATATACAT GAGCACTGTT CTGCTCAGAG CTCCTGATCT ACCCGACCC 540  
30 CTAGATCCA GCACTGGTTC AAGCTGCTAT GAACACGCC CCGTGGACA MACCTGGAA 600  
TGGCTGGAG TGGAGAAA CTTGAATTC TCTTTCCCTC TCCCTCTTCC AAGATTACTG 660  
35 GAACTGTATC CTATTAGAT CTTCGAGCT TGTTTCCCTG CTGGTGGGA CAGAGACAA 720  
AGGAGAGG AGGCTCTAGA AGAGGAGCC CTTCCTTGTG CTCTGGGTA AATGAGCTTG 780  
40 AACTAGATA AATGGAGGA CAAAGACCT CTGATTTTAA ATTTCTAATA AATGTTAGAA 840  
GTATATATAT AATATATAT ATTTCTTAA ATTTTTCAGT CTGTGATAG TCTAAAAATC 900  
CAATTCCTCT CCGCTGAGC CTGATGAGA CACATGAGA AACTGTGTTT TCAATTAAG 960  
45 ATGTTAATA AATGTTGAA ACTTGAAAA AAAAAAAAA AAA 1003

## (2) INFORMATION FOR SEQ ID NO: 163:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 163:

60 AAGAGCGGC AAGAGGATGT CGATTCTTAC ACGAGAGTGG GAGAGATTAAC CAGCGACTTG 60

GGGAACATC AGCATATGCA TGACCGAGAT GACTCTATG CTGACGAGAT GGAACGAGA 120  
5 ATGAGCGACA AACTGAAAC AGCTTTTAA ATTTCAATG ACAAAGTAGA GGCCTTAAT 180  
AAGAGAAC TGGAATTTGA AGTCCCTTTT AGGCACTTGG GATTTAAGCG AGCTGCTAT 240  
AGAGTAACT GCTTCCTTCA CCGCACTAGT AGTGGCTGG TTAAATCTTAC GGAATGGCA 300  
10 CTTTTTGTGG TGACATTTGA TGAGTATAG CTATCCACT TTAGCGGGT CCAATTTTAC 360  
CTGAGAACT TTGATATGGT ATCTCTCTAC AAGCACTACA GCAAGAAAGT GACCATATC 420  
15 AACCCATTC CTGTAGCTTC TCTTGACCCC ATCAGGAAT GTTGAATTC CTGCGACTG 480  
AATACACAG AAGGATAGA GTCCCTCAC TGACTTAAA TCAATGAAGC CATTTGTAT 540  
GACCTGAGG GCTTCTTCCA ACAAGTGGC TGGTCTTTCC TGAGGCTCA GGGTGAAGG 600  
20 AGTATGCTG AAGAGGGGA TTGAGATCT GAAATTTAAG ATGAGACTTT TAATCTTTCA 660  
CAGATGACT ATGAGAGGA AGAGAGGAC AGTGAATGAG ATTATTCAT ACAAAGCAGA 720  
GACTGAACT ATTCTAAGA GTCTTGGGT AGTGAAGAG AGATGGAAG GAAATGGAT 780  
GACTGAGG AAGAGCCCG AAGAGCGAC CAGAAAGTC GTTACGAGA ACAAAGAGA 840  
900 CAAAGTCCA GTATGAGCG GAGAGGAGG GATCTGTGAC AGATTTGGG CCGTGGCTCT  
30 AAGCTGGTT CAGAGCAG CTCTGACC CCGAGAAA AGAGGAAGTA ACTTCTGAC 960  
TTTGGCCCTG AGCTCAATTC TTCTTCCAGC CAACCCCTGA AATTTTACA TGACATAGA 1020  
35 ACTATATTT TCTTTGGTT TTCAATTTGA GTTTTGGAT TTTGTATAT GGGTTTAAAG 1080  
GGCATTTGT GTGAGCAAT CTACTGGGG AATTCAGGC CACAGAGAC AGTGGCAAT 1140  
GGCCCATTC AATGGCAG GAGAGGCTG TTCTTGAAG CAGGAGAGG CTCCCGCTGT 1200  
1260 TAATATATAT TTTTCTATC TTCTCTCTTC CTGTGACTT CTGCGAGAC ATTTGATGCT  
1320 TCTGACTCT TAATTTGGT CTGAAGCTG TAATTTGAG ACGATGGTAC AAGGTGACCC  
1380 TTAAATACC GTATCATGGT TCTTGACGAG CACATTCAT CTTCCAACT ACCCTACTGC  
1440 CATGACTTC CGCATCTC TAAGTTTTAT CTTTGCATA CTCAGGTTT TCGGAATTTT 1500  
GCTAATGTT GTGTAACCC ATACAGCTTG ACCCATGAG GCAATTTGG CTGGTGCCTT 1560  
50 CTTCTGAGT TTCTGCTTT CTGCGCTGT GCAATTTCTG AGGTATATCT GCTGCGCTTG 1620  
AAGCAATAG AAGCAATAG ATCTCTGCG TGGTTATTT TTCTCTTACA ATGCAACAT 1680  
GTTAATAGA TAGAGGCA AATTCCTACT GTCTTCTTTT TTTTCTCTTA TATCTAAGA 1740  
55 AGATATACA GGTGTGCTT CATGTACCG TTCTATGAA ATGTAGAGA AGGTCMAAG  
GAGTAACT TTAGATCTG AAGGAGAG TCAATGCTTG GGCATGAAT ACCCTATGA 1800  
60 GAAAGAGA GAGAGGGA GGCATATCT ACAAACAA CCTCTGCGA CTGCTGCTCC 1860

TTATTTTAACT TTCTCTCTTC ATTTCTCTCT ATTTATCACA GTTCTCTCTT AACAGCTTT 1920  
CAGTATTTTGG GGGAGTTTAT TTCTCCATCC TCCCTTTCTG GTTCTCTGCA CCGACCTCTC 1980  
CCACTCCAGT TCTCTCTCTG CTCTCTGACT TTATAGAGAG AAGCGGGGAG GGTCTCCCGA 2040  
TTTTATGTTT GTTGTGTTT TTCTCTTAC AGTAGAGCTT GATATTTTCA ATTTTGAG 2100  
10 AACTAAAGA TCAATAACT GGTGTTTTT TGTGTCTCT TTTGTAAA AAAAAAAAAA 2160  
AAAAAAAAA AAAAAAAAAA AAAAAA 2196

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1945 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 164:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 164:

GCACAGATC GCGCGAGCG ACAGGAGAG GAGGAGGCG GGTCTCCCG GCGCGCTAC 60  
CCAGAAAGCA CCGAGAGCA GACAGAGTG GGTCTCTCC GAGCGAGCC CCGAGGAGC 120  
CCCCCGCCCG CCGCGAGAG AGCGCTTTC CAGCGAGCC GACTCTTAG AGGAGGAGG 180  
CGCGAAGAG AGCTCAAGC TCACACAGC CCGTCCGCC AGCGCGGCCA CTCGAGGCT 240  
CTCTCGGACT CCGCGGCTC TCTCTGAGT CTCTGAGGAG ACCGCTCTG CAGAGCCAT 300  
GGAGTTGCTG CTGTCTCTCC TCCTGAGGCT GCTGGCGCCC ATGCTCTCTG CCACTCCAGC 360  
TGAAAGAGAG AAGGAATGG ACCCTTTTCA TTATGATTA CAGAGCCCTG GATTTGCGGG 420  
ACTGTGCTTC GCTGTGCTCC TCTCTCTGTT TGGGATCTTC CTATCTCTAA GTCCGAGTG 480  
CAAGTCAGT TTTATCTAGA AGCGCGGCG CCGAGGAGT GAGGAGGCC AGCTGAGAA 540  
CTCTATACCC GCGATTCGA CAGAGGCCA GAAAGCAGAG AACTGAGTG CAGCCATCAG 600  
GTGGAAGCTT CTGAACTTG AGCGGCTCG TTGAACTTTT GATTCGAAAT GTGATCTCTT 660  
AAGAAAGCG GCGACTTTC CAGAGCCCT TTCCCGAGGA GAGCGAAGA ACTTGTGTGT 720  
CCCCAGCCCT ATCCCTCTCA ACACATTC TCACCTCAT GATTCGAACTA ACATCTCTCT 780  
CCCCACTGCA GCGTCCGCTC CTGCGGACT CCGTCAATGT GTGTGTGTGT GTGTGTGTGT 840  
GTGACTGTGT GTTGTGCTA ACTGTGCTCT TTGTGCTTAC TTGTGTGTGT ATGTGTGTGT 900  
GTTTGTGTGT GAACTGTGTA CTGCTTTTC CAGCGAGGG CTGAGCCACA TGCGGCTG 960  
CTCTCTCTG CCGCGGCGC CTCCTATCAC CTCTCTCTC TAGGAGGCTG CTGTCTCTC 1020

GAGACAGCC CCGTCCCTG ATTTAGGAT GGTAGGGTA AGAGCAGGG CAGTGTCTT 1080  
CAGTGTCTT GGGACTGGG AAGGTTTCA GCACTTTCT ATCATCTTC ATGACTCTT 1140  
5 TTACTCTCTT TAACAAAGC CTCTCTCTT TATCCACTT GATCCAGTC TGAAGTCTC 1200  
TTAGAGCTG GAGATCAA GCAAGGACT GGTAGGCCA GGTTCAGCT CAGCGAGCT 1260  
ATGCGCTCC GTGTATAT TTCTCCAGG GCTTCCAGG AGGATTCCT ATCTCCCG 1320  
10 CCGCTTACA GAGCGCCCG GATTCAGG CCGAGGCTT CTACTCTCC CTCTCGAAT 1380  
GTGTCCCTG CATCTCTCT CAGCATTAC TCTATGAGT CTGGAGCTT ACCCTTTCA 1440  
15 ACCTTCCCTG CTCTGAGAC TTCAATCTAC AGCGCAGCT ATCCAGATC AGACTACAT 1500  
CCCTCATTT GGTCTCTCT CAGCGATTA TTAAGACT CTCTCTCTT TCGGCGAGC 1560  
ACACCGGAT GATGAGAGG AGAGCAGG CTTTCTCTT TCTGCTAG TCCCTTTAG 1620  
20 TCGGAGCAG AGCGACTCC CCGATCTTT CTCTCTCTG TCTGTGCTA GAGCGTGA 1680  
CAGGTGCTT TCGAGCTCA GCGAGCTCC TCGAGGCTT GCGAGAGTG AGAGTTTGA 1740  
25 GGTCTATAG AGATTCGAA CTCAACCGC ATCCGCGCC TCTCTCTCT TGTGTCTCG 1800  
CGAAACCAA CCAAGCCCTG GCTGTGACC CATCTCTT CTCTGTATG TATCTATCC 1860  
TCACAGCAA CAGAAAGAG GATTAATTA TCTTTTCTT CTATGCGAA AAAAAAAAA 1920  
30 AAAAAAAAA AAAAAAAA CTCA 1945

(2) INFORMATION FOR SEQ ID NO: 165:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2933 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 165:

GGTTCAGCC AGCGTTCGG CAGCGCTCT TTGATCTCT GCTGCGCGTG CCGCTCTCG 60  
GATCAGAGC CAGTGTATAC GCGCGGCCA CCGCTTGTGT GCGCTAGAG GAAAGAGAA 120  
50 GAGGCGCGC TCGGTCTCT GCGCGAGCT CCGCTCTCT GCTGAGAGC CAGCGCGCG 180  
CCAGTCTGT CCGTCTCAC CCGTCTGAG CCGTCTGAG CCGCGCGAC AGCGCGCGC 240  
CGGAGAGGC GCGCGCGAG GCTTCTGAG CCGATTCAA AGTCTATAC CTATCAGAG 300  
55 CAGTCTCAA ACAGAGAGC CCGCGATC GGTATATCT TATGAGGC ATCATAGAG 360  
ACAGATGT GTGTCTCTG TCGAGGCCA AGATGATCA ATTCAGTGT TTCCGAGTG 420  
60 ACAGATGT GTTCAGAGA AGAGAGAG CAGAGCTCT TTGATCTCT CTCTCTGAG 480

5 ATACTTCTTC TCAATGAGAG ATTGCGATCA ATAGAGTTTGT TCGGATTTAC CTTCGTGTAC 540  
GCTTAGGGGA TGTATTCAG ATCCAGCAAT GCGCTGATGT GAGTATCGGC AAGCGTATCC 600  
ATGTGCTGCC CATTGATGAC ACHETGAGAG GCHTACTGCG TATCTCTCTC GAGGTATACC 660  
TTAAAGCTCTA CTTCCTGTGAA GGTATCTGAC CCAATCTGAA AGGAGACATT TTCTCTGTCC 720  
10 GTGTGTGGAT GGTCTGTGTG GAGTCTAAGG TGTGTGAAGC AGATCTCTAG CCTATATGCA 780  
TTTGTGTCTC AGACACAGTG ATCTACTGCG AAGGGAGGCC TATCTAAGCA GAGGATGAG 840  
15 AAGATCTCTT GATATGATTA GGTATGATG ACHTGTGTGT CTGCGAGGAG CAGCTAGCTC 900  
AGATAAGGGA GATGTGTGAA CTGCGCTCTA GACATCTCTC CCTCTTTAAG GCAATGTGTG 960  
TGAAAGCTCT TAGAGGATTC CTGCTGTGAG GACCTCTGTG AAGAGGAGAG ACCCTGATTTG 1020  
CTCGAGCTGT AGCAATATGAG ACTGAGGCTT TCTCTCTCTT GATCTATGTT CCTGAGATCA 1080  
20 TGAGCAATTT GGTCTGTGTG TGTGAGGCA ACCTTCTTAA AGCCTTTGTG GAGGCTGTGA 1140  
AGATGTCTCT TGCATCTATC TTCTATGATG AGCTAGATGC CATGCTCTCC AAGAGAGAGA 1200  
25 AAGACTATGG CAGGTGTGAG GCGGCTATGT TATCTAGTTT GTTGACCTTC ATGTATGGCC 1260  
TAAAGCAGAG GGCACATGTT ATTATATGAG CAGCAACCAA CAGACCCAAC AGCATTGAGC 1320  
30 CAGCTCTTAG GCAATTTGTT GCTTTGACA GCGAGGTAGA TATTTGAAAT CCTGTATCTA 1380  
CAGAGCCTTT AGAGATCTTT CAGATCTATA CCAAGAAAT GAGCTGTGCA GATGATGTGG 1440  
1500 ACCTGTGAACA GTAGCAATG AGATCTAGGG GCAATGTGTT GCTGACTTAG CAGCCTGTGT 1500  
35 CTGAGAGGCT GCTCTGTGAG CCAATCTGAA GAGATGTGAT CTCTATGACC TAGAGATGTA 1560  
GACCATTTAT GCGAGGTCTA TGAATCTCTT AGCATTTACT ATGATATGACT TCCGZTGGCC 1620  
40 CTGTAGCCAG AGTAAGCCAT CAGCTCTGCG GCAAAACCTTG GTAGAGGTGC CAGAGGTATC 1680  
CTGTGAGGAC ATCGGGGGCC TAGAGATGT CAAAGCTGTG GTAGAGGAGC TGTCTCTGATTA 1740  
45 TCTCTGTGAG CAGCCAGACA AATCTCTGAA GTTTGTGATG AGACTTTCTA AGGAGATTTT 1800  
GTTCATATGA CTTCTGTGCT GTGCGAAAGC TTTGTGTGGC AAGCCCTTG CTATGATATG 1860  
CGAGGCCAAC TTCTATCTCA TCAAGGTGTC TGAAGTCTCT ACATATGTGT TTGCGGAGTC 1920  
50 TGAAGCCAT GTCAAGGAAA TCTTTGAAA GCGCGCCCAA GCTGCGCCCT GTGTGTATTT 1980  
CTTTGATGAG CTGATTTGCA TTCCGAGGCC TGTGTGAGGT AACATTTGAG ATGTGTGTGG 2040  
GCTGTCTGAC GATATATCA ACGATTTCTT GACAGAAATG GATGTGATGT CCAAGAAAAA 2100  
55 AATATGTGTC ATCATTTGCG CTAGCAAGCG GCTGTGATTC ATTGATCTCTG CCAATCTGAG 2160  
ACGTGGCCCT CTGTATGAG TATATATAT CCAATTTCTT GATGAGATGT CCGCTGTGTC 2220  
60 CATCTCTAAG GCTAAGCTCTC GCAATCTGCC AGTTGCGAAG GATGTGACT TGTAGTTCTT 2280

5 GCTTAAATAT ACTAATGCTT TCTGTGAGC TGACTGTACA GAGATTTGTC AGCTGCTTG 2140  
CAAGTGGCC ATCTGTATAT CATTGAGAG TGAATTTAGG CAGAGAGCAG AAGAGCGAGAC 2400  
AAACCAATCA GCAATGAGG TAGAGAGCA TGAATCCAGTG CCTGAGATCC GTTCAGATCA 2460  
CTTTGAGGAA GCAATGCTT TTGCGCGCC TTCTGTGATGT GACATATACA TTCTGAAATTA 2520  
10 TGAATGTGTT GCGCCAGACC TTCAAGAGAG TCGGGGCTTT GCGAGCTTCA GATTCCTTTC 2580  
AGGAGACAG GGTGTAGCTG GCGCCATCTA GCGGAGTGA GCGGZGACAG GTGTGAGTGT 2640  
15 ATACACAGAA GACATATATG ATGACCTGTA TGCCTAAGTG GTGTGTGCCA GCTGTGATGT 2700  
AGCTGCTCTG CCTGTAGCTT GTTCTCTTGG GTTGTGGGGCG CTTCGCCAGG AAGAGAGACA 2760  
GGGTGTGGCC CACAGGCTGC TCAATCTCTC AGTCTGAAAC GTTCAAGCTAC AGTCTGACTC 2820  
20 TCGAGAGGGG GTTCTCTGTT GAAATATCA AAGCAAAAGC GATTAATATTA AAGCATTTT 2880  
CATTTGTATA AAAAAAAAAA AAAAAAAAAA CCGCGCGGGG GCGCGAAACA TTT 2933

## (2) INFORMATION FOR SEQ ID NO: 166:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2243 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 166:

TCGAGAGCC GCGCGGGGNG GCGCTCTTGG CAGGAAAGCG CTTCTGTGAC GGTGTATACC 60  
GATGCCAGAA AGTGGCCTTG GCGTGGGGAT CACCATAGCT TTCTAGCTTA GCTGTATCNC 120  
GCATTTCTTC GTGTATATG GTTCTATATCA GTATATACATCC CCAATTTTCC TCTATATTCG 180  
TTCTTGGCTC CTTTGTATAT TTCTTCTAGG AGCGTCTACG GTGGGAGCA TAGGAGGACA 240  
45 GTTATGCTATG GGTGTGTCTG AAGAGCCCCA TAGTGATTTCA GTCTTCAAAA CCAACCAATC 300  
TGAAGCAGAG GAGATTTTGG GAGAAATATC TGAATCTGGA TTATGACAAA GATTTATCTT 360  
50 TTCTTTATGT AATCTATTTTA GATCGGGCTG ACTGTACAAA TGAATCTCTG AAAAACTCT 420  
TCACTAGTTC TAGATATGGG AGGTGTGAAA TGAATCTTTA CCTGTGAATC TTCTCTTTCAC 480  
TCCCGGCACT GCGCGCTGTC TGTGCGCTGG AGCATTTCTC CCAAGCTAGC TGGZTTTCAG 540  
55 CAGTGTGGAG CTTCGCCAGT ATTCGATTTTC ATTCTATGTA TTAAGAAAG TTGCCATATTT 600  
TCAAAAGCTT GAACTATGAC TCAATATACA ACCCGGAGTT TTGTGTGATG GCGCAAGAGA 660  
60 GGTATGTGTA TGTGTCTTAA GAACTATGTA GTATGTGTGTA ATAGAAATTA TATTTATCCA 720

5 AAGATTTT AAAATAGGG CTGTGTTTAA AAAAAAACC AAAACAGGAA AACAGGCT 780  
GATTATAG AGGTCAACT CTAAATGGGG TCGGGGGCTG GCAAGCTTC ACCTCAGC 840  
TCCTCCCTCC TCGATGGGG TGTTTACAGT GTACACAGTG TGTGTATCAG CATGTGTTCA 900  
ACTGCTGTGC ATTCCTCCG TGGCAATTTG TGTAGCAAT CTACTAGAC AAAAGGCAT 960  
GAAAGTCTT GGTTCGCACA CTGCGATATA TTGGATTTT CACTCAGTT TATGAGTTT 1020  
ATTTCGAAT CCATGATCAT CTAAAGATGA ATACCTGTCT GCAATGTAT TCAATCTTAG 1080  
TGAGCCAAA TTTGTTGTTT GTTACTACAG AATAGAGATG ACTGTTTTT GCAACAGCC 1140  
TATGATTT GCAATCTGTG ATTCCTTGT AAAAGAGAGA GTGCATATG CACTGCATTA 1200  
AAGTGTGT GTTCTATGTC AATGATATG GTAGGAGAA TGTATTCAT TAAATGGATA 1260  
GACATACCA GACCTAATTT GCAAGTATG GGTCTTAAAC TTCAAGTCCA ATGTATATCA 1320  
AAACCATCT GAGCTGTGA TCTCTTAAT ATTATTTT TTTAAGTGT GAGATGTG 1380  
AGAGAGGT CTCATTCAT TTCAAGTCTG CCTGAGGAA ACTCGGCAT GATTTCTTTC 1440  
AGTGTGAG TTCTTTGT GTTACACCT CCACTGACC CTCACCTTC GAATATCC 1500  
AGTTTGTG GTTTGTGAT TTATCTTAT AATTTACT TTTGTATTT TGCATTTAC 1560  
ATGTGTTG TTTGTTTAA ATCTGTGAA AGTGGTTGA TTAAGACT CTTTAAAT 1620  
GGAAGCAC AGTCAGAGA ATGAGAGCT AGAGAGCT GCTGTGAC GCTGTCTTT 1680  
GTGTTGTT TGTGATGA AGATCTTG CTGGGTTTT TTGCTTTGT TTGAGGAAA 1740  
TGTCTGGAG TAAATTTAA GTTCTGGAG TTAATTTGT TTACAGGAT TTTGTTTT 1800  
AAAAATAG GATCTCTG AACTTTGAA TCAGCCCTT AATATTTTC TGAATGAA 1860  
AACATTA CAAGAAAT TTCAATGAA GATGTGAGA TTTATGAAA AACAGAGT 1920  
TATTAATGA AGCAGGAG TGAATCTTTA AAGAGACT GTTCAGGAC ACGATGAG 1980  
TCTTTCTC CGAAGCCGA AGTAAATCTA TATCTGTTAG AATATATGA GCCAAAGA 2040  
TGTAAATTT AGATTTTT TTCCATAGT TTAAGAGAAA TATATGACC AAGTGATTT 2100  
GAGTTTAA AATGTAAA TATGATGAC AAAATTTGA CTCATACAGA TTGAGATC 2160  
TAGTAGGTT CACTGATA CTAAATTTT ACATGTGT TCTTTTGA TCTATTTTT 2220  
ACTTTATTA AGGTTGAAA ACC 2243

(1) INFORMATION FOR SEQ ID NO: 167:

(A) LENGTH: 1816 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(E) SEQUENCE DESCRIPTION: SEQ ID NO: 167:

5 GTGCGAGAC TTATTAATTC CCTTACAGG GCGCTATTA GCGGAACCT TCCCGAATT 60  
TTGCGGTGAA CCGACGGTC CCGCAGGCT AGAGAGGAA GTTGTATGTC CAGAGGTGA 120  
GCGAGAGGC GCGATTTCT GCGGGTGAAG GCGGAGCTG AATGTACAC GAGAGCGGAA 180  
GCAAGGTGTC GTGGGCGGA GAAAGGGGCT GCGCCGAGA GAAAGGAAA ACCCTTCCA 240  
GAAACAGCA ACAGCTGAG CTGCTGTAC AGAGGGGAC AAGTGGGCG CCGCGAAGGG 300  
GAGGCTCTG GTAGAGACC AACTGGGCT CCGCGCGCT CTGCTGTGTA CCATGGCTT 360  
GCGCGAGCT TCGGGAGCG CTTCGGCTGA AGCATTTGAC TCGGTCTTGG GTGATAGGC 420  
GTCTTGCCAC CCGGCTGTG AGTTGAGCTA CCGCTTGAC ACCTTACCTA AGGAGAGAA 480  
GTGTAGCCA TGTACAGAG GTTCCAGCT GTTTTCATTT TGTGATTTG TGAATATGG 540  
AATGACTTA AATGAACTA AATGGATG TGAATCTGA TGTACAGAG CATATTTCCA 600  
ATCTGATGAG CAAATGCTT GCAATCTTG ATCCAGAT CAGCTGCCAT TCGCTGACT 660  
GAGAGAGAA CAATTTAT CTCTGATCC AAAAATGAC CTACTCTTTC CTCTAATCT 720  
GCTGAGCTA TTCTGAGTG AGATGATGA CTCGCGAG AGCTTCTTAA CTTCTGTAT 780  
GACTTTTAT CTTCAGGCG ATGACGGAAA AATGATTTA TTCTGTCTA AGCCAGAAA 840  
TCCAGGTA GCAACATTT TGAAGCGAG AGCCCTTACA AATTTGAGG RAACTCTCT 900  
AGCAAAATG TCTTCAAT CTGAAATGAG AAATTCAAA GCGACAGAA ATTCTCTGA 960  
ACATGAGAA AGTGAATCT TTTTAAGATG CTCTCTCTT AACTGTGGT GATTTTAC 1020  
TACAACTCT GTCTCTGCG TGAATGATTT GCTTTGGAT TGTGTGTGAA CTGTGTGCTA 1080  
CAGCTGTG GACCAATTT AGTTTCCCT TGAAGCTG AATATGATG GTGACTTGA 1140  
GTTTATGAT GAAACAAAG TAAACAGTA TCGAGCTTCT TCTCTGTG TTTGTATGTC 1200  
TAAACTGAA GATCTGAG AGCAGGCG TCTACTTACA AATGTGATC TTGCTGATTC 1260  
TGAATTTAA GCAATTTTCT TTTAAAGAC AATGTATATA GACATCTAAA ATTCCACTC 1320  
TCAATAGCT TTTAAATG TTTGATGGA TATAGGCTT AAGAAATCAC TATAAATGC 1380  
AATTAAGTT ACTCAATCT GTGAAAAA AAAAAAAA AAAAAAAC TCGAGGCGG 1440  
GCGCTGAC AATTCGCTT ATATGATGAT GTATTTTAT TTTACTATA TCTGTAGCTA 1500  
TTTGTTTTT KCTTTGTTT ATGTTTTT TCCCTTCTT MAGCTATGAG CTGATCTATG 1560  
CTGCTGCTA CTTCTGCTA TGTACTGTC AATTACTTA GTTACAGGC TGAATATTTA 1620  
GTAGAAATG TGTCTCTCT CAGGATGCG CCAAAATCT GTAAATTTGA ATTATGAGG 1680

AAATGACCTT TATGACACT AGATTTCAG GACTGAAAT CATTAATTT TTATTGAAAT 1740  
 AATTATGTC TCAAAAAA AAAAAAAA AMMRARASK RRMACTGCA GGGGGGGCC 1800  
 GGTACCAAT TCGCCG 1816

10

(2) INFORMATION FOR SEQ ID NO: 168:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 945 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 168:

AGAAACGTT GATGGGACTG AGAAACGGA GTTAAACCT CTTTGAGCT TCTGAGACT 60  
 CAGCTGAAAC CAGGGGGGAC AGTTGGGAC ACCATCACT TCTCCGAGC AGAGAAACC 120  
 GAACCCACCA ACCAGGGCCA GATAGCTG AGAAACATC TACAGCGGA AATCAAGTT 180  
 ATTGGACTA TCGAATCTT GTTGGGATG ATGATATGA GCTTGGGAT CATTTTGGCA 240  
 TCTCTTCTCT TCTCTCCAAA TTTTACCGA GTGACTCTTA CACTGTTGAA CTCTGCTTAC 300  
 CACTATPAG GACCTTTT TTTTATCAT TCTGGCTCTC TATCATATC CAGAGAGAA 360  
 AGATTACCA AGCTTTTGGT GATAGAGC CTGGTTGAAA GCATTTCTGAG TCTCTCTGCT 420  
 GCTCTGTCGT GTTTATCAT CTTGCTGTC AGAACGGCA CTTTAAATCC TCCCTCATG 480  
 CAGTGTGAGT TCGACAAAA TATATACCA ACAGAAATT ATGTTTCTTA CTTTATCAT 540  
 GATTCACTTT ATACCAAGGA CTGTATACA GCGAAAGCA GTCTGGCTGG AMCTCTCTCT 600  
 CTGATCTGTA TTTTCACTCT GCTGGAAATC TGCTTAGCTG TGCTCACTGC TGTCTGGCG 660  
 TCGAAACGG CTTACTCTGA CTCTCCGCG AGTGTACTTT TCTGGCTCA CAGTTACAT 720  
 GGTAAATCTG GCATGCTCTC AAAATGACT CATGACTGTG GATATGAGA ACTATTGACT 780  
 TCTTAAAGA AAAGGAGAA ATATATACA GAAGTTGAT TCTATATGA ATATGAGAAA 840  
 GTTAAACAT ATAGAAAGC AAAGTTGAG TTTCTTAAT GTAGCTTTT AAAGTATCA 900  
 ACATAAAAA AAACATAT TTAATCTGA TTTAAAGATA ATGTC 945

55

(2) INFORMATION FOR SEQ ID NO: 169:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 902 base pairs  
 (B) TYPE: nucleic acid

60

(C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 169:

GGGAGACCA CAGAGAGAT GAGGAGACC AGGCTCTGG GGTCTCTGT GATCTCTTT 60  
 GTCTCAGAC TCGAGCTGC AACTAATTA ACTGAGAAA AGTATGAACT GAAGAGGGG 120  
 CAGACCTCG ATGTGAATG TGACTACAG CTAGAGAGT TTCCGAGAG CAGAAAGCT 180  
 TCGCAGATA TANGGAGGG AGAGTGGCC AGAGCCCTGG CATCCAGAG GAGGCTTTCA 240  
 AGAAATGCC ATCCAGTCCA AGTGGGAGG ATCACTACG AAGACTACCA TCAATCATGT 300  
 TTACTGCGCG TCGAATGCT CAACCTTCA GTGGAAGATT CTGACTCTTA TCACTGTGTG 360  
 ATCTACGAG CTCCAGAGA GCTTCAGAT GTTGTGATC GCATCCGCTT GCTGTGACC 420  
 AAGGTTTTT CAGGGAGCC TGGCTCCAT GAGATTCTA CCCAGATGT GTATAGATT 480  
 CCTCTACCA CCACTAAGGC CTTGTGCCA CTCTATACA GCTCCAGAC TGTGACCCA 540  
 GCTCCAGCA AGTCACTGC GATGTCTCC ACTCTGACT CTGAATCAA CCTTACAAAT 600  
 GTGACAGTA TCACTAGGT TCGGTGTTC AAGATTTCA TTCTCTGCG TGTGTGATTC 660  
 CTGATTAAGA GCTGTCTCT CTCTGTCTG TTTGCTGTA GCTGAGCTC ATTGTACCC 720  
 TAGGCGGCG AACCGAGG AATGTCTCT GACTTCAGC CAGATCCATC TGGGATGT 780  
 GCGAAGGAG GAGGAGGAG GTAAAGGCA GGAATTAAT AACTGATTA AATCTCTTAA 840  
 TACCGCTTA AAAAAAAA AAAAAAACH GAACTTGG TTTTCACTC CATCAGCTCC 900  
 TT 902

(2) INFORMATION FOR SEQ ID NO: 170:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1883 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 170:

AGAAACGAC TGAANAACA CATTTTCTA CATACAGCTG GCGAGGTAGC TGAAGACTG 60  
 GCACTGGCA CACATCTAG GTTGAGAG AGTTAGGAA ACCAGAGGC CAATGGATC 120  
 TCGTGGAAA CCTGAACCT GTCTCTGGG CTCTCTTAC AGTTCTGAG TTGAAATTC 180  
 TTTTATGCC TAGCATCTG TTGATTTTA ACCCGAGG CAGCCATGTC ATAGACTAT 240  
 GTTACTCTT GTTTTACTT TGTTTTAAT GTTCTTAGA CCGAAGTGC TCCCTCTTT 300

5 TCCTCTCTTG TGGTACCTG TGGCCTCATG GAGCTCATG CCCACAGCTTT CCCACTTTCA  
GCAGTCTCTTT GCTCTCTTTG CTCTACCTC AATATAGCCC AGGAGTGGGC TTATAGTCTCC  
10 ANATAGGAGC ATTTGAGCTT TCTCTGGGG GATGGGGATT GGGATGGGCA GATCTGTTT  
TGGTCTCGG GGTATTTCC AGTGGTGTG AAGGAGAGC TGGGCTTTTC GCTCTCTAT  
CCCTGAGGCT GGTATAGAG GACTGTATCT ACACCTGTTTC TTCTCTACCT TCTCTTTTGT  
15 TAGGAGGACC TCATCTTAG TTCTCTAGA GAGTCTCTGG CTTAAGAGCTG TAGGAGGCT  
GTCTTAGGCT GGGATTTGG AGGAAAGCG TCGATAGGCG ATGATATAGG TATGAGATGG  
GCCTGCAAAA TCAGACAGAA ATGGCTTAG AAGCGGAGG GAGCAGTCC TGTCTCTAG  
TCATAGAGTA TGGAGGGAC CTCCTTAGCT TGGAAATGA GATTTGAGG GGTATAGAC  
20 AATAGAGTG CTTAGTTGAG GATGTCCCA AAGTTTGTGC CATCTTATC ATTAGTAGT  
TTTATAGACC ACAGAGGCA ACCGAAAGG GAATATGTT ACTTTGAGG CTTATTTT  
TTGTCTAGG TGTGCTTGG TACATGACA AGATGCTAT ATGCTGACA TTTTGGCTTT  
25 AAGTCTTAG GACTTTCCC ATTATAGCT AATGGAGGA TACAGATGG CAAGTCTGCT  
TTTGTGTTT TTTTATAT TTTTGTGTT TTGCTCTGTC TTATGAGAT TTTTACGAT  
GCAGAGAGT GGGAGAGTG GTCTTGGAC CCCATGTGC CATGACTAG CTCGATCACT  
30 TATGAGCTAT GGTCAAGCTG GTTTTATCTG TATCTCTGTC TTTTCACTG TATTTTAT  
TCAAAATCA AGACACTTG CCATGCGAC GGTCACTAT CATATACAA TCAAGTCTGC TTTTGGCTTT  
35 TTTGATGTC ATAGTAGTGG GTGTGACTAT CATATACAA TCAAGTCTGC TTTTGGCTTT  
TATTTTATC TATATAGTT CAGAGATGG GCTTTGAAA TGTGTTTTTA GATTTAGCA  
GGAGTCTCA AAGAAATGA GAGGATGCT TCCTTTCCC TTGCATCTAC AATACAGAG  
40 AGAGACTGTT CTGTTGAAA ACTCTTTCA AATTTCTGAT ATGATAGGT ACTTGAGCC  
CTTACACGA ATGTCAATCT TTTTCTCTGT GTACATGGA AACTGTGTC ACCATTAGCA  
45 TTGTTATCAG CTGTACTGG TCTATACT CTGTTTTGG AGATATAT TGGATTTGT  
TGTGTGTC TGTGAATA ACCTGCCCA AATATATG AACTGTTTT TCTACTTGT  
AATTTGAC CTTGTTATA ACCGATAT TCTGTGTC TTAAAGATA TAAATGTTG  
50 TAAATTGCA TGCATGTC AAAAAAAA ACCGTATCT CTGTTAAAA AAAAAAAA  
AAAAAAAAAAAAAAAAAAA

55

(2) INFORMATION FOR SEQ ID NO: 171:

(1) SEQUENCE CHARACTERISTICS:

60

- (A) LENGTH: 2100 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 171:

5 TACTTTTACA TTACTGCGCT TCMAAAAGTG CTTATTTCTGA GCACATATA CTTTATTCCT 60  
10 TACATATGTA TGTACACAG GTACCCAGAG TGTACTGTG GCAGCTTTCA AATACATAC 120  
ATCAGAAACA GTAGTGTCTG AGATAGGGA ACTTTGCCA ATGTAGAAA GTCACTACT 180  
TCCATATGCT CCTCTTCAG CGGTACCTT GAAAGGGCT GCMAAGACAT TCCCTAGCA 240  
TCCCTTGTCT ATACAGCTTC TTATATTTTA TATCTACTG GATGTTAGCA TATGCTTAG 300  
GTTTCTCTGA CTCTGCTTCA AGGAAATGTA AGTTTATGCG CATGAAACA TTTAGAAAA 360  
AATAAGATCT TTAAGAGAT TATAGAGCC GTAGTCTGTA TTAGATGTG TGTCAATAT 420  
GTGTTCTATA AACTATAGAT CGTGGGTTT AGATGTTTAA ATGTGACGA CATCTCTCT 480  
CCTTTTGTCT CTCAGCTAA CATGAGAA ATAGAGAAAG TCTTGGCTGT GGGATTTGGA 540  
AGCTCAGGG GCGAATGTC CTTCGAGT CTTTAGACA TTACTTTGAC TCTTAAATAT 600  
ACTATGTTAT GTTATTTGAT GCTTTTGT TCCATAGTTC CATCACTGAC AATACTGTCA 660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560  
1620  
1680  
1740  
1800  
1860  
1883



400

5 GATGAGCTGC TACTCTGANT ACANTCTCCT GACTTAGGGA AAGCAGACGA AAAAAATATA 1680  
AAGAAATATG CTATTTTCTAT ATGTGTGATA CTGAGAGAGC CATGTGATTC CTAAATATATA 1740  
GGTTTCTCTT TTITCTGTGA TTCTTAGGGA ATTGCATTTA TTCACTACAT TACAAACCT 1800  
CACTGATGTA TCCAAATATG CACACATATG TCAATATGAA AATAGAGGAA TAAATATCTGT 1860  
TATAGCAGAG TCAATATATG ATTTTCTTTT GTGTATATGC ATTATCTGAC TATATATATA 1920  
CCTGTATTTT TATTTACCTT CTATCATGTT TCTCTACGAA TTATGTTTTT TCAATCTCTT 1980  
ATAGAGTGA ATATGAGAT TATATTTCTT TTITCTGTAA AAGATTTGCA ACTACTTAT 2040  
TATATTTAGA AATCATATTA ACTTCTTAT ACATTTTAAA AAAAAAATA AAAAACTGGA 2100

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1930 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 172:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 172:

30 CATTGATATG TGCTGCGGG TCGAAATGCG CAGCGCTCC GCGCGGCTC GTGTGTTGCC 60  
CGCCATGCGA CTGTGCGGG GCTTGCGGG GAGCTGGCT GAGCGGGCG CGCGGGGCG 120  
GCTGCTGTG GTGGGGGCG CGGCATGCG CTGCGACTC CTCAGATATC TGTGCTGAC 180  
CGTTTCTCC CACATCGACC TGAATGATCT GATATATAT GATGTAGCA ACCTCAGAG 240  
ACAGTTTTTG TTTCAAAGA AACTGTGTGG AAGTCAAG GCAAGGTTG CCAAGGAGAG 300  
TGTACTGCGA TTATACCGA AACTATATAT GTTGCCTAC CATGACAGCA TCATGAAACC 360  
TGCTATATAT GTGAATTTT TCGACAGTT TTTACTGTT TTTACTGTT TTTACTGTT 420  
AGCTGCGCGA AACATGTTA ATAGATATG CTGCGAGCT GATGTTCTC TTATGTAGAG 480  
TGGACAGCT GGTATCTTG GACAGTATAC TACTATGAA AAGGTGTGA CCGAGTGTTA 540  
TGAGTGTAT CTTAACCGA CCAAGAGC GTTCTGCG TGTACATATC GTACACACC 600  
TTGAGACCT ATACATGCA TGTGTTGGG AAGTACTTG TTCAACAGT TGTGTTGGGA 660  
AAGAGATGCT GATCAGAG TATCTCTGA CAGACTGAC CTGAGACTG CTGCGAGCC 720  
AACGAGGCC GAGCGGAG CTAGAGATC TATGAGAT GTTGACATTA AACTATATTC 780  
TACTAGGAA TGGCTAAT CACTGATTA TATCAGAT TATCTGAT TATCTGAT 840  
TAAAGATAC ATCAGTATC TGTGAGAT GAGAACTA TGGCGAAA GGAACCTCC 900

401

5 AATTGCTGTC ATATGGGATA AGATGAGCC ATCTGCAATG GATTTTTGCA CCTCTGCTGC 960  
AAGAAATGGA CCCAGTTAG GCTTGAAGA CAGCAGGTT CTAGATGTTA AGAGCTATGC 1020  
AGTCTTTTTT TCAAGAGCA TCGAGACTTT GAGGTTCTAT TTACAGAAA AGCGGATGCG 1080  
AGCTGAGCTC ATATGGGATA AGATGAGCC ATCTGCAATG GATTTTTGCA CCTCTGCTGC 1140  
AACTCTGAG ATGCAATATTT TCAATATGTA TATGAGAT AGATTTGATA TCAATCATAT 1200  
GCGAGGAGC ATATATCTCT CTATGCTAC TACTATGCA GTAAATGCTG GGTGTATAGT 1260  
ATTGAGGGA TTGAGATTT TTTCAGGAAA ATAGAGCAG TCGAGAGCA TTTTITTTGAA 1320  
TAAAGAGCA AACCCAGAA AGAGCTTCT TGTGCTTGT GCACTGATC CTCCAGACC 1380  
CAATGTTAT GTATGTGCA CAGGCCAGA GTTGACTGT GGTGTGATG TCTATATAGT 1440  
GACTGTCTC ACTTACAG AGAGATAT GAAAGAAA TTCTCTATG TAGCAGAGA 1500  
TGTCAATAT GAGATGGA AAGAGCAAT CTATATAT TCCGAGAGG GAGAGAGGA 1560  
AGTATATAT CACAGAAAT TGTCAATAT TGAATTAGA AATGCGACC GGTCTCAGC 1620  
AGATGACTTC CTCAGGACT ATACTTTAT GATCAGATC CTTCATATG AAGACTAGG 1680  
AAGAGCTTT GATTTGAG TTGTTGTGA TCCCGCGAA AAGTGGGG CCAAGCAGC 1740  
TGAAGTCTC GCGAAGCA TAAACATG GCAATGATG TGGAGCTTC AGCCCTCCAC 1800  
CTTCAGCTC TGAAGAGCC AAGATGAGC TTTCATATG TTGATCGG TGAAGAGAT 1860  
TCTCAATTA TTGCGGAGC TTCATTTAG GAGAGAGG GAGGCCGCC AAGAGAGGA 1920  
TTTAGATATG 1930

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 173:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 173:

50 GCGCTGCCC TCTGCTGTA GGTGCTTAG GACTTCGGG TGGCTCTAAG GGGAGGAGAT 60  
AGGCTGCGG AGCGCGGCC TGTGCCCCG ACCAGCCCT TCTGTGCGG GTTCACCCC 120  
GATCGAGTG GTACATGCT TACGCGGGA CAGTACTG ACCGACTCT TCTCTCAGCA 180  
CCTCATGTC GTGCTGTCT CTCTGAGG CAGGCTCG CCGAGCCCTG TTGACAGGA 240  
CTTCTACTCC GATTTTGGG AAGAGCAC AGGAGAGAT GAGAGTAC AGCTGATCA 300  
CTTATGCCC GTCAAGTTTA CTTAACCCAG TCGAGAGG AGTGGGACC TGAAGTTAC 360

5 TGTGGCCCAA AAGATGGCTG AGCAGAGAA GGGCCCGAGCC CTCAGCATCC TGCCTTAGCT  
GCAGGGCTTC CAGGTGGGCA TGGCAACCCC TGGGTGGCTGC AGGGGGCCCC TGGCCCCCAA  
GACACTCTTG CTCACCGACT CCGAGATCTT CTCTCTGGAT GAGGACTGTG TCCACTACCC  
ACTGCCGAG TTTCGCAAG AGCGCGGCA GAGAGACAGG TACCGGCTGG ACATGGGCG  
CGCGCTCCGG GACCTGGACC GATGTGCTAT GGGCTACGAG ACGTACCGCC AGCCTTCACC  
CTCTCTCTTG ATGAGCTGCA AGCTATGAC CTCATGGGCA GTGTCAACCT GGAGCACTTT  
GGGAGGTGC CAGGTGGCCC GGTAGAGCC AGCGAGGCC GTGAGTCCA GTTGGAGTGG  
15 TTGTGCCCCA GTCTGAGAG CAGAGAGAG CTCATCTGCC TTGTGGCTGG CCAATGGGAG  
GGCTCTGTG GCGGTAGCT GCTGTGTGAG CTCACCGCT AGCCAGGCC ACAAGCAGCC  
900 TGTCTGTGCC AGCCTGAGCC CTACTGGGGC AGGGACAGAG GCTTTTGTGT TCTCTAATAA  
20 TGTCTGTGCC TGGCTTGTGT ACTTATATT GACTGTCTC GAGAGAGATG TGACATGTG  
TGTGTGTGT GTTATTTCTT TCTCATTTG GAGTGAGAA TGGCGGGGCC CTCAGGGCTG  
1080 TGGGTGTCT GTACGCTCC CAGCATGTGT ACAGCGTGC ACACAGATGT GGTGTCTGCT  
TGTGTGGAC GGTGTATAC AGGTGAGACT GTGGGTCTGA CTTTCTCTTC TACAGTCTCT  
1200 TTCTGAGT GTGAGTCCA GTCTTGTGT GGTGTGTG TTGCTGTGCT TGTGTCTGTT  
1260 GGCATCTTGC TGTATCTCT GAGGTGTGTA GCGATGCA CATTGGAGC TCCGACCCCA  
1320 TATTTGCTT CAAATGTGAG GTCTCTCCG ATCGAGCAA GTTGGAGAG CCGTGGGGGC  
1380 AGGGAGCTG GAGCTGGCAG CACCAAGCT GATTCCTGCT GCTGTATTC TCTATTCGAA  
1440 TAAAGCAGG TTGACACG TCAGAGAGG AAAAAAAAAA AAAAAAAAAA ATTTCCTGGG  
1500 CCGCAGGG

45 (2) INFORMATION FOR SEQ ID NO: 174:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3173 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 174:

55 TGCAGCCCAA GCGTCCTGC TTGTCACAG AGGTTAGAC CTCGAAGAG ATGGCTGAG  
ACCACTATG GATCTGCTC CTTTGGCTGC AACTGGCC GAGAGAGCT GGAAGAGCT  
60 CAGAAATCTT CACAGTGAAT GGGATCTGG GAGATCACT CACTTTCTCT GTAAATATCC  
180

AGAGACCG GCAATTTAAA ATCATTTCTT GCACTTCTAA AACATCTGTT GCTTATTTAA 240  
CACCAAGAA CTCAGAAACA GCAACCGTAG TTACTGTGAC CACAGAAAT TATTTATGAC 300  
5 GATACATGCG CTTAGGTCCG AACTACATC TGTGTATTAG GATCTGTAGG ATGGAGAGCG 360  
CAGGAGACTA CAAAGCAGAC ATAAATACAC AGGCTGATCC CTACAGCACG ACCAGGCGCT 420  
AGAACTGCA AATCTATGCT GCGTTGGGA AACCAAAAT TACACAGAT TTAAATGCGAT 480  
10 CTGTGAGCAG CACTGTAAAT GTACACTGTA CATGCTCTGT AGGAGAGAA GAAAGAGATG 540  
TGACATACA TTGGATTTCC CTGGAGAG AGGGTAAAT GCTTCAATC TTCCAGACTC 600  
15 CTGAGGACA AGGCTGACT TACAGTGTTA CAGCCAGAA CCGTGTGAC AGCATTTCTG 660  
ACTCATGTC TGGCGGGCAG CTCGTGTGAG ACATGCAAT GGGCTTCTCT ACTCAGACA 720  
CGGGTGTCT GAGGTGCTG GCTATGTTCT TTCTGCTTGT TGTCTATCTG TCTTCACTGT 780  
20 TTGTGTTGCG TTGTGTTGAG AGAGACAG ATCTGCTTC AAGAGAAAC ATATACACAT 840  
AATCATGCG TTCAAGGAC ACCGAGCAG CAGATCTGAG AATCTATGAT GAATCTCTGC 900  
25 AGTCCAGGT GCTTCCCTCC AAGAGAGAC CAGTCAACAC AGTTTATTTCC GAAATGCAAT 960  
TTGCTGATA GATGGGAAA GCGAGCAGC AGAGACTTAA ACCTCTGCGG ACTTCAGCT 1020  
ATGAAATGCT GATCTAGGCT GCTGGGCTGA ATTCTGCTTC TGGAACTGA GTTACAGCA 1080  
CCAACTAGCG CAGGTTCCTT GATTCAGAT CTTTCTCTGC CAACTCTTAC TGGAGATG 1140  
1200 CAAACTGCA CATCTGACC TGTAGCAA GCAAGAAAC TTCTGCTGCG CATAGCTGT 1260  
35 GCTTAATGCG ACAAATGAT GATAGCTT CTTGAAATGA CTCCTTCTG AATGATGAC 1320  
AAGCAGGTT ACTTAGTATA GTTTTCCAA ACTTCTTCCC ATCATACAC ATGTAGAAA 1380  
TATATTTTTT ATGGCACT GCGTAAACA AGCAGATG CTCAGTCTG GAGCTGCTAT 1440  
ATGACTAGAG GCTCTCTG ACTGGAGTA ACAGCCCTGC CCAATTAAGT TGGAGAGG 1500  
GATCAATAT TTTCACACCC TGTATAGCG CATGCAAC CAGCGAGAT GCTCTGCTCA 1560  
45 CAGTCAATAT GTGTGAGAT CCGTGTGCG TGGGCTTTC CAGCGATCTT GAGCAATTA 1620  
GGAAATGTA CCGTTCGCTT GAGGCAATG CAGCCCTTCC CCGAGTCCA TGGCTTGGAG 1680  
AGCAGATGT GGGCTGATA TAGCACTCT CATTCCTTTG TCTGGGATC TTGTGTGAG 1740  
50 GCGTACAGG CTTAGTAGT CCAACACAG ATGACATGCG TGTGTGGTTC TCTGTGAG 1800  
TTGTGCTCT CAGCGATGTA GACAGCTCT CCAATGAG TGTGAGAAA TGTCTTCTCT 1860  
55 GCAAGGTGTA GAGATGCTG GAGACTTTT CTTGGAGTGC TACTTCAAAA GCTTTTAGG 1920  
ATTTCCTTC TGGCAGAT TTCTCTGCT ATCACTCAA GCAAGCTCAG CAGAGAGC 1980  
AGCCATGCC AGTATTTCCA CTCGCAAA GAGACTGACC AGCTTATATT TCTCACTCT 2040

CTGGGAACT GGTATTAATC CAACTATCA ATTAGAGAC CTTCAGAGA GCAGATCAT 2040  
TCTCAGAG GAATCTGGA GATGATGTT CAGATGATCA AACTGGGTTT ATCCAGTTT 2100  
5 CAAAGACTCA GAGACTAGA GTTTAGCTG AGCCAGAGTG CCGCAGCTT GCGATGCCC 2160  
AAGAGCAT CAGCAGCAG CTTCAGAGG CATTAATCTT CCTCATAGAG GAGATATAT 2220  
TCAACTGTA GCGAGCATTT CATATGATCA TTTTAGGAG TGTTCCTCTT ATGTGTTAGC 2280  
AAGTATATC GGTAACTCC TAAATCCAA TGAATAGTCC TAGCTGAGC AGCAATGGCC 2340  
TGCATTAGG CAGTATAGA CATCACTGCC AATTAATGAA TCCATAGACT CATTCAGAC 2400  
15 CAACTAACAT TAGCACTATG TTAGAGCTG CAGGCCCCA AAGTAGAGA TGTCTATAT 2460  
GTCTGCTCTT GTTAGCTCA GAGACATTT CAGCAGACA CACTACATTT AACCTGAC 2520  
TCACTGTC AATTAATGCA TGAACATCA GAAATATAC TTATGAGGG GCGAGGCTCA 2580  
AGCTGGGCTT TGAAGTGG ATCAATTTG GATAGAGAT GAGCAGACA GAGGCGCTCC 2640  
AATGAGAGA AGCATGAAA ATGAGAGGG GCTTGATCA GTGGGGTGTG TTGAGAGC 2700  
25 CTCTCCAGAT GCACATGCA TGGTCACAT CCGTTGCTTA TCTGTGGCAG AGTGTCCAG 2760  
CGAGATGAT GCGCCGACC CATGTCCAT TACATGCTT TCAATGCTT TCAATGCTT 2820  
TACTCTTCT GTTAAGCTTT CCGTGTATC AGAATCATA ATTATCAGG GATCTTTTCA 2880  
CACTGCTTT TTTTCTCTT TGTCTCTTT ATCACTAAA CTCATCTCAT TCAGCTTAC 2940  
AGCATAACTA ATTAATTTT TTCTCACTA CATGTACAT GTGGAAATTA CAGATTAAG 3000  
35 GAGCGCTCT GCGGTGTGG CTACGCTTG TAACTGAC TAACTGAC ACTTTGGAG GCGAGGCG 3060  
GCGATCACC TGAAGTCAG AATTCAGAT TATCTGCCC AACATGTTGA AACCCCATTT 3120  
NFACTAAAA TACGAATTTA CCGAGGTGTT GTGCGACCA TCTGTAGTCC CAG 3173

## (2) INFORMATION FOR SEQ ID NO: 175:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 991 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 175:

AAATTCGCA CAGCTGAG GAGAGAGAG GAGAGGCGG CAGGACCAA GTGAGAGCA 60  
TGAATTACA GTGTGTTTC GTTGGCTCC TGGGTACAT ACTGATATG TCGTCATAG 120  
AAGACAGG TCTGAGAGA TGTGTATTT CAGAGCAT GACCATATA GAGAGATTT 180  
60 TCCAGAAAT CAAAGAGCC ATCCAGCTA AGCAGACCTT CCGAATGTC ACTATCTGT 240

CCACATTCGA GACTCTGAG ATCATTAAG CCTTACATGT GTGCTGCTG ACCAAGAAC 300  
TCTTGCGTTT CTACGTGAC AGGTGTTTCA AGATGATCA GAGCGAAGC CCGAATATCT 360  
5 TGAAGAAAT CAGCAGCATT CCGAACTGTT TCTCTACAT CAGAAAACT CTGCGGCAT 420  
GTGAGACA GAGCGATGTT CACTGAGGC AGGAGGCCAC GATTCGACC AGATGATATC 480  
10 ATGACAACTA TGAATAGCTG GAGTCCAGG CTGTGCCAT TAAATGCTG GAGAGCTGG 540  
AGTCTTTCTT AGCTGGAAT ATTAGAGATC ATGAGATAT GTCTCTAGT TGAATGAG 600  
GAACCTGAT ATGATGCG GATGAGAC CCGCTGTGG GTTACTGTT GAGAGAGCC 660  
15 CAGCTTAGAG GCGAGGAGA TGGGAGGCC CCGTTCGAG TGAAGTCCC ACTGCTGCG 720  
CTCAGGCTCT CTATTTGCG TTGAATATG CCAAAAGTC TACTGTGTA TTTGTATATA 780  
20 ACTCATCTG CTGAAGGCC CTGAGGCCA TCTGTGAGT AAAGGCTGCT CTTCGCAAT 840  
AATTTATGTT GAATGATAT ATTCATGTC TGTGATGCA CCGAATGAT ATCTGTATG 900  
25 ACACATGTA CTGATGCTT TTTCTGATA ATTTCAATAT TTTACTTAAA AAAAAAAAA 960  
AAAACTGCA GGGGGGCC GTACCAATTT T 991

## (2) INFORMATION FOR SEQ ID NO: 176:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 176:

ACAGCGCTCT TGGAGGCTG AGCCGCGCTC TCTCACTCA CTTCAACCC CAGGGGGGCC 60  
CTCACAGGG CCGCTCTCTT CCGTGCAGG CTCTGCTGTT CTGCGCTGCT CCGTGCAGAG 120  
45 AACAGGCCA TGGTTCGCG CCGTCTGCTG CCGCTGCTG YCGTCTGCTG GCGCGCAGA 180  
TTTTCGAC CTGTGCTG CACAGATCT GGTCCAGCT ACTTTATGG GTTCACTCAA 240  
CGAAGACC TTCAAGCTC CATGGTGGC TCTGTGGAAA TCCCTTTCT CTTCATATAC 300  
50 CCGTGGAT TAGCAAGY TCCGACCTG AGAATATCT GAGAGCGGG CCACTTCCAC 360  
GGGAGTCTT TCTAGAC AGGCGCGCT TCCATTCACA AGAATATAT GAACCGGCTC 420  
55 TTTTCAAT GAGAGAGG TCAAGAGAG GGTTCCTCA GATCTCAA CCGTGCAGAG 480  
GAGAGCAT CTGTATTTT CTGCGATC GAGTGCACA CCGGAGATC AGGAGGCGAG 540  
60 CAGTGTGAT CCATAGGG GACCAACTC ACATCAGCC AGGCTGTGAC AACCAACCC 600

5 ACCTGAGGC CAGGACGAC AACCAACATA GCGGGCTCTA GGGTCAACA AGCAAAAGGG 660  
CACTCAGAT CATGCACTT AAGTCTGAC ACTGCATCA GGGTTGCATTT GGCCTGTGCT 720  
GTGCTCAAA CTGTCTATTT GGGACTGCTG TGCTCTCTCC TCTGTGTGTGG AGGAGAGGA 780  
AAGTACAG GCGCCAGGC AGTGACTTCT GACCAACAGA GTGTGGGGAG AAGGCAATGTG 840  
TATTAGCCC GAGGAGGTS ATGTGAGCC GCTTTGTGAG TCTCTCAGC TGTTTTCCCA 900  
TTGGCAGAT ACATGAGAG CACCTGAGG ACCTTTAATA GGCNAAGCG CAGGCAAGA 960  
GAGGCTGGG TCCCTGATC ACCGACTGGA GAGAGTTAC CTACAGAGC CTTCATCCAG 1020  
GAGATCCAC ACTGCATCA TATAGGAGTS AGTCTGAGC TCCACTGAT TAAACCACTG 1080  
GCATTTGGG GCTGTTATTT ATAGAGTGC AAGAGTTCC TTTATCTCTC CAGAGATGG 1140  
AAATACAT TTAATTTGCT TACCATGAC CCGTTATCTC GTGCTCCACA TTTTCCATC 1200  
TGTATGTGCT CTGTCTCTA TCCGAGAGG TTTTGGGAA TAAATAGCT GAAATGATC 1260  
TCACTAATA AAAAAAAAAA AAAAATCGA 1290

## (2) INFORMATION FOR SEQ ID NO: 177:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2290 base pairs  
(B) TYPE: nucleic acid  
(C) STRAIN/GENESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 177:

5 TGGGGCCCC TTTGGATGCT CTGGGTGTTT TTGCGAGAG TTACAGGATG TCAGTGTGG 60  
GGAGCTCAC ACCCTTGCTG TCGACCAATG AAGGCTGTTC CAGACCAAGT GCTTCAGAC 120  
ATTTCAGGC TCCAGGAG AGGCTGGGAG CCCCCACAGA AAGCAAGA AAATGCAAA 180  
AAAAACAT CTTTTATTTT TTTTCTCTT TTATTATGA AACAAACAA ATGCCCCAG 240  
AGAGGGTCC ATGATACCA GAACATCCA AAGATACATTT CTACCATTTT TATCTGTGG 300  
TGTGAGGCC AGCATGACA TAAACAGCT AAATCTACTA CATTTGACT ATTTTCACTA 360  
ACTGACATTT ACAGATAT ACTAGAACG GCATTAATA GTTTAGAAA AGTTACGCTA 420  
AATCTGATG CACATCATC AGAAGATGA CATTTTAAT ATAAAAAGA AAAATCTCT 480  
GGAAGCATTA TCCAGTATTT AAGGACAGT GCTACTCTGG ATGTGACAAA TTCTGTATGT 540  
GGCTGTACT CTTTCCAAA AGCATCTGAG AGGCTGAGT GCTGCNAAG AACACACCA 600  
AAACACCA CACAAAAA TGTGTCTTAC AGTTGTTAAG CAGATGACA CTGCCCCACA 660  
CAAGAGGAG TCTGGAGTTC AGTTACGCC GAGGCTGCG CCGCTGGCC TCCAGGGTTC 720

5 ATTGAGATG TTCTCAATC CAAITTCGAC ACAGACTTG TCACTACTGC TCTGCTCTTG 780  
AAAAAGCAT GTTAGAGCT GCGTACAGG TCTCAGGAGT GGAACAACTT AATTGAATCA 840  
CGCAGCTTT CTATACAGA AGAAAGCGAC GTGACTGTCA CCGTCAAGCC GCGAGCAGG 900  
GCGGTAGGA AGTCATTAAT CTTCTGAAC TCTGAAMGA AACCATGTGT GAGTCTCGA 960  
CAAAAGCTT TAAANAAGT ACAGACCAA TCGAGTGTCT CAGTTATACC NCGTGGCT 1020  
GTCAAGCTCA GTGCTTCTT TCTGAATGA AGAGCAGAG GCGAGCGGAC GCGACCTCA 1080  
CAGAGCA GCGCAGAGG AAGGCCCCA CAGTCTGCG TGTGCGCTGC CCGCAGGCC 1140  
GGCGCTGCC CCGAGGCTC ACHAGGCGAG CAGCTCACTG CCGTGTGCT GAGAGGGCAT 1200  
TCCAGAGAG GCGCCGCGC CCGAGGACC CCGGCTTAG GGTGTACGTA TCGAGGCC 1260  
CTGTCTGCG AGCAGCTTAC CAGCAGGCT GCGTGAAGC CTGTCACTG TCGTGTGTGA 1320  
ATTGCTTAA TTGGTTTAA ATAGTCCATT AAGTCTGT TTGAATTA CTTTGAATA 1380  
CGAGGTAC TTTAATAAT GGAACCTTC AATCCATTT ATATTTTAT TATTAACAA 1440  
ACTAATTA AAGTTTACA AACTGCTCA AACTCACA AGTGTGAG TCAACGCA 1500  
TTTAAAAAT GATATTTAC CAGCATCTC TCTACAGT TCGCTCTCA GTAAAGGTAT 1560  
ACCTACAT GTAGGCTCA GTGACTCTG AATCAATTT ATGCTTTT TAAATACGC 1620  
GTGTATGAG ATACTATGA TAGTCCAT ATCCATCCAG AATCTGCG AAAAACTCT 1680  
GCGCACTTA CAGGACAGC CAGACAGAG CTCCATACC AGCTCTGCC TCGAGGTGCG 1740  
CGCAGCTGA GTTCCGGAG AATGCTGCT TTCACTAT TCGGACTTAA CTGTACAC 1800  
CGTTCAGAG GAGCAGCTT GCGAGTCTC GGAAGGAT CCGCTGCGG CCGCAGACT 1860  
CTTCACCCC TGGGAGGCG AGACAGGCTC GGAAGGCT GCGCAGGCA CTGAGGCTG 1920  
CGAGGAGCA GCAATGTCCA CCGCAGGCC TCGAGGCTA ACTGTGCA TCGTGGCGG 1980  
AGCGGCTAT CTGATGTGTS GCGAGTTTG GGAATGCGG GTACTCAAG ACCATATGC 2040  
AGCGCTGCG AAAAAAAC CAAACATCA CTTCTATTA AACTCTGAT AATTATATT 2100  
TTTACATAG AAGTTAATA ATCAGACTT AGATTTACTA TACATTTTTT CTCTCAGAT 2160  
ACAAAGTTA TATTATATA CTGGGTTC CTAAATGAT TTCTTTTAAA ACAGTCTTAA 2220  
AGAGACCA AGTGAATCA AAGACATA ACAAATATA AATTAGAT GTCTGTAGC 2280  
TGAAGCTGT 2290

(2) INFORMATION FOR SEQ ID NO: 178:

60

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 549 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 178:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

GGCAGAGCC ATGCGTGGCC TGTCTGTGAT TCTTACAGTC ACTTGTGTGG CTGTGTCGTA  
CTGAGCAGCT ACTGCAATG TGGCAGAGG ATGACTATTT CTTTCTCAGG AGGGCAGAAA  
TGTGGAATAG TGTCTGTCCA TGGCTCTCTT CATGGCTAC CAGCTCTGCC ACCTGTGTTA  
ATCATATACA ACCGAGAGAG AGCTGCTGCG AACTGACTTC TGGGAACTCC CTGGGATGCT  
TTGTGTCAGG AATGTATAGT GATATACCT GTTGTGCTGG ATCTGGGGCC TGTGTGATGT  
AGTAGAGAGG TAAAGGGCCA CATTCTCTTT GACTCTGTGG GAACTCATTC ACAAAGAGAA  
TGTATTCAGG ATGCTTCTGA AGATTGGCTA AAATATAGCG GTTTTCACCC CCGTGTATCC  
ATCATCTCTA GATCTCTCTT TGCAGAGAC CAGAGAACTG ATTTACAGAA GTGACATGAA  
AACTATTCAT CCGAGATTTT GATATAGTC AATTATAGTT TCTAGCTATT AAAGAGAAA  
AAAAAAAA

(2) INFORMATION FOR SEQ ID NO: 179:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1509 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 179:

40  
45  
50  
55  
60

GGCAGAGGG CTCATTATTT CCGCGCGGG CCTGCGAGAC ACTTGCGGCC TTCTGAGCC  
GCGCGCGGCA TCGCGCGCGG CAGCGCGCGG CATGTGCGGC CAGAGCTTCG AGACGCCCTC  
CGCATTCGAG ATCTGCGGGA TCTATGCGGC AGATGATGCC AACGTGCGCG GCAATGTCCA  
CGGGGGGAGC ATCTGTAGAA TGTATAGAA GCGAGCGGCC ATCATAGCA CCGCGCATTT  
CAGACCGGAG AACGGGAGG CTTGTGTGCG CCGCTGTGCT CTTGTGTGAG CAGACGAGTT  
CCTGTCTCCC ATGTGATGCG GTGAGTGTGC GCAATGTGAG GCGAGATCA CTTACAGCTC  
CAGGCTCTCT GTGAGGTGTC AGTCTAGCT GATGTGCGAA AAGATCTTCA CAGGTGCGAA  
AAAGCTGACC AATAGAGCCA CCGTGTGTTA TGTGCGCGCT TCGCTGAGAA ATGTGAGAA  
GCTCTCTGAG GTGCTCTCTG TTGTGTATTC CCGGAGAGAG CAGGAGAGG AGGGCGGAA  
GCGGTATGAA GCGGTAGAG TGGAGCGCAT GAGAGACGAG TCGAGGAGG GCGACATCTG

650  
700  
750  
800  
850  
900  
950  
1000  
1050  
1100  
1150  
1200  
1250  
1300  
1350  
1400  
1450  
1500  
1509

CGACCAAGTC CTCAGCCAG AGCGAGACAC TGTCACTAC AGCGATCCA CTTTGTATCA  
CTGTGCGGG CTTTCAAGCT GCACTCTGCA CGCTTTTGTG CAGGAGGTG TGAACATGAA  
GCTCATGAT GAGGTGCGCG GCACTGTGCG TGCAGGCCAC TGCAGACCA ACATCTCTAC  
AGCTTCTGTG GAGCGCATTA ATTTCTGTA CAGATGAGA AAGGCTGCG TCAATACAT  
CTGCGAGCC ATGACTTTCA CGAGCAATTA GTCAATGAG ATCGAGGTGT TGTGTAGCC  
GAGCGCTTT GTGAGAGCT CTGAGAGCG CTAGCGGGCC GCAATGCTCT TCTTCACTTA  
CTGTGCTG AGCGAGAG GCACTGTGCT CACTGTGCCC CAGCTGTGTC CCGAGAGCCA  
GAGAGAGAG AAGCGTTTGT AGAGAGGCA AGGCGGTGAC CTGAGATGA AGGCGAAGCR  
ACAGGGCAC GCGAGCTTC AGCGGTAGAC TCGCTCTGTC TCGACTGCT GCTTCAAGTA  
GCAATGCGAA CGGCGCACT GTCACTCAC TTAGAGATTC CCGCTTGTGC CAAGAGCCA  
ATTCAATTT AGAGCTGTG TTGTCTGAG TTGTGTATTC ACAGTGTGTA CTTGTACTCT  
CTCTGCGAA CTTACAGCC AAGCTTTTAT TTATATCAT TCAATATCAA TCTTACAGAG  
TGTGTGCGG AGCGCGGGA GCGTTGCGC AGAAGCCCTC GCGAATGCTT CCGAGAGCC  
TGTAGGCTAT GCGAGAGCC CAGAGCACT AATAGAGCTG CTCTTGTGCT GAGAGAGAAA  
AAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA  
AGAAAAAN

(2) INFORMATION FOR SEQ ID NO: 180:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1316 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 180:

60  
120  
180  
240  
300  
360  
420

AGCTATATCA TAGGAAGAT GCGCAGACG CGGTACACAG TAAATGTCTC TCGGGCGAG  
CGAACCCAG TCGCGGGGG GCGCTCAGCC TAAATTTCCAG CGCCAGGCC AGCAGCGGCT  
GCGCTCGCG TTCTCGCTGC GCGTCCAGCC TGCATCTCTA GACCTCTCGG CAGCAGCGCC  
TGAAGTGGG GCTCTGTGCC AGAGCCCGGC CTCAGCGCA ATCAGAGCA GACCCGAGCG  
CGCGCTGCG CTGTGCTGCG TCTTTCAGGG CCGTTCGCGG CGGGCGCGCT GTTCAAGCTG  
CAGCGATCA TTGTGCGCT CATTTGTGAC AGCTAGAGAA CAGCGAGCA GCGTCTCTCC  
CGAATATCG GCACTCTGTT GCGACTGTC GAGAACTCT CAGTGTAGGT CACCAATTC

TTTTTGGTC GGCAGATCA CTCAGAGAT GAGTGGCTG TTGACATGA ATTTCCTAG 480  
AAATGTAAG AACTGATTA AAAGTTTCT CCAATGAG TCATCTCTGG CTGTACCT 540  
ACGGCCATG ACATCAGCA GACTCTCTG CTGATCAT GACTACTACA CCGCAGGC 600  
CCCAACCCC ATCAGCTCA CTGTGACAC AAGTCTCCAG AACGGCCCA TGAACATCA 660  
AGCTTACCT AGCACTTTTA TGGAGTCCC TGGAGGACC ATGGAGTGA TTTTCAACC 720  
TCTGACGTC AATTAACCT ACTAGCAC TGAAGCAT CCACTTGACC TCATCATGA 780  
GACTGCTTT AGCCCAACA GAGTATGCT ACTCTCAGT GACTTCCAG CAGTAGAGG 840  
GGCATCAGCT CCGATCCAGG ATGCCCTGAG TACAGTCTTG CATATCCAG AGCATCTACT 900  
GTCTGAAAG GTCTCAGCT ACATGACT GTGGCCCTTC CTGATGACC TGGTTACCA 960  
AGTACCGAAA ATAGTCTCCG ATGACTTTGA GACCATGCTC AACAGCAACA TCATGACT 1020  
TTTGTGCTG ACTTACTCG CCACTCTCAC ACATCTCAG ATTGCATCA ATGAAACT 1080  
TGTAACTCT TGAATGACC CCAAGACTA CACTTCTCG TCTAGTATTT AACCCAGCA 1140  
CTCAGAGTG AAGGAAAT GGGTTTTTG TGGTCTTGA TCACACTGAG ATAGTCAAT 1200  
GTGTGTACT CTATTAAGG GAGCTTACT TTTGTAAAT TAAAAAATA AAAAAACN 1260  
SGGGGGGGG CCGGTCCCA TTSSCCCTTT TGTATCTGT ATTAAATCC CAGGCG 1316

## (2) INFORMATION FOR SEQ ID NO: 181:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 777 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 181:

GCATGKCA GACATGACTT CTATGCCAG GCTGTCCAG TGGCAGGGTC ATCAGGCA 60  
CATCATTAG GGTGCTCTT ATGCTCCCTG CTCTGGATC CACAGCGGG CTATCTGCT 120  
TTATGGGCT GGGACTAGA ATTGATGCT TCAAAACAT CACTGTGTGG CCAACAGTT 180  
TGACCAAGG GTAGATGATA ATGCTCTCA GTGCTTAGA GAATACTTAC GTTATAGGG 240  
CAATCTATT GGGACTGAA CTTTGAAGC CACATATG AAGAGGCTT GCTTACCTGT 300  
TGGGGGCCAA GAGGATGTT ACCAAATG GTYAGGAA YTTGGYKGG AMCARRRGG 360  
GKGGGARRH CRGGGYTGG SCATTTCSK KGGWACCTT TTAGGTTAAR RGGGCKGTW 420  
ATTAGATTGT GGTAAAGTA GCACTTTTG CCGTGCAMA TTTGCTGCTT GGTGAAATGY 480  
TCTTGTGTC TTCTACCC CTACCTTAG TATGTTCTCC ACTTACTTTC TCATTAATG 540

## (2) INFORMATION FOR SEQ ID NO: 182:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 182:

GGCAGATA ACTATGACA TGTATGCTT AAATATTTT TTAAATTTT TATCTGCG 60  
ACTGCTCTC AATGCTGAC ATGTGTCCA GGGAGCAAT GCTTCTTGT TTCTGAAT 120  
GCTCTTTAG ACTGCTCTT TTTCCCATCT TCTCAGCTCC TGCCCTCTCT TCAGGCTACT 180  
TCCGTGCCA GAACCTCC AGGTGAGAG CAGAGAGAA GCTCTATGG TCACAGCAC 240  
AGATGGGC TGGAGATCA TTCAATGCT TTTGGCTTGA ATTTTCTGTA TGGTTTACT 300  
GATCTGGGA AGAATATC TTGCCAGAA AATGATAGN CTTGACAT OTTGATGAT 360  
CCTGACAC CTTGAAGAC ATTCTTAATA TGGTTTGTCA GGCAAATGG TTATGATCA 420  
TTTGTGCTT GAGGTAGAG TCTTGAGAA TCACAGACT TCATGATTA AATGCTGACT 480  
TGCCCTTGA CTGGCTCTG TGAATGCG CTCTGACT GTTCCAGTA GATGTGACA 540  
CACACACT ACAGGACA COTGTGGGC TGTGACTAG GGTCAAGCT CCGTCAAGC 600  
CACTATGAG AATCAGCTT TTACATGGG CTGTTTCTA CTGTTCTAT GAATCTGAT 660  
TGGTTCTTT CCTCATACC CTTCTGCTT TTCAATGTTT TGTGTTAGTT TTCTGCTTT 720  
TTATTTTAA CTGAAATA AATCTTTAG GCTATTCACC TTAGCTTATG AAAAAAATA 780  
AAAAAAACT C 791

## (2) INFORMATION FOR SEQ ID NO: 183:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1405 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

AGATAGAAA CTCTGTGAT AGGAGAGTG AAGAGGGAAT ATGTGGTAGA GCACTTAT 600  
TCAGTGAAT GCTCTGCTT AGCTTTTCCA TTCTGTGAG CTGGCTTCC TAAATATTC 660  
AGGTTTGTA GCTTGAGCA GAACTTTGAT GGAAGAGAA CTTTCCCTTC TGTACTGTA 720  
ACTTAAAT AATAGCTCC TCAITCAAG TAAAAAAA AAAAAAAA AAAAAA 777

## (2) INFORMATION FOR SEQ ID NO: 183:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

GGCAGATA ACTATGACA TGTATGCTT AAATATTTT TTAAATTTT TATCTGCG 60  
ACTGCTCTC AATGCTGAC ATGTGTCCA GGGAGCAAT GCTTCTTGT TTCTGAAT 120  
GCTCTTTAG ACTGCTCTT TTTCCCATCT TCTCAGCTCC TGCCCTCTCT TCAGGCTACT 180  
TCCGTGCCA GAACCTCC AGGTGAGAG CAGAGAGAA GCTCTATGG TCACAGCAC 240  
AGATGGGC TGGAGATCA TTCAATGCT TTTGGCTTGA ATTTTCTGTA TGGTTTACT 300  
GATCTGGGA AGAATATC TTGCCAGAA AATGATAGN CTTGACAT OTTGATGAT 360  
CCTGACAC CTTGAAGAC ATTCTTAATA TGGTTTGTCA GGCAAATGG TTATGATCA 420  
TTTGTGCTT GAGGTAGAG TCTTGAGAA TCACAGACT TCATGATTA AATGCTGACT 480  
TGCCCTTGA CTGGCTCTG TGAATGCG CTCTGACT GTTCCAGTA GATGTGACA 540  
CACACACT ACAGGACA COTGTGGGC TGTGACTAG GGTCAAGCT CCGTCAAGC 600  
CACTATGAG AATCAGCTT TTACATGGG CTGTTTCTA CTGTTCTAT GAATCTGAT 660  
TGGTTCTTT CCTCATACC CTTCTGCTT TTCAATGTTT TGTGTTAGTT TTCTGCTTT 720  
TTATTTTAA CTGAAATA AATCTTTAG GCTATTCACC TTAGCTTATG AAAAAAATA 780  
AAAAAAACT C 791

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 183:

5 AATTGATTA ACGCTTGA AGAAGCTCT GCTTTGAG GCTTAGTAG CAGCACTGCC 60  
AGTAGCTGG AGCTGGAGA ACTTGGCAT GAGAAAGGA TCGAGGGA GGAATATAG 120  
AAGCTATGG GCGAGATACA TTAGCTGGA TCGAATATC AGAATATGA GCGACAGAA 180  
GTTAATGAG CAGATCAGC AAGAGAGAG TTACAGGATC TCGATGACC ATAGCTGGG 240  
CAGAAAGAT CAAACAGAA ACTAGAGCA GACTGGAGC GACTGAGCA GGAATTTCCAC 300  
TATATAGAG AAGATCTTTA TCGAGAG AAGCATATTC AAGCAGAT TAAATATGA 360  
GACGAGAAA TTCAAAACT CAGGATCAG CTTACCAATA AACTTTAG CATTACAGT 420  
CAGCTGAGT TAGAATATG ACTGATCAG CTAAACAGGA CTCTCATCA GAAACAGAC 480  
20 ATCTGTGGA GTCTCAGAC AGAAGAGAC TCCCTGGTCT TTCAACTGA GCGGCTGGA 540  
CAGCAGTGA ACTCCGCTC TCGAGTAGT AATATGGGT CTTCGATTA TATCTCTGA 600  
ATTGACATG GTTAGGACAC TGTCTGCA AATGTTCCTG TCTTTTAAA TGAACAGAA 660  
ACTAATCTG CAGGATATGA GCGAAAGTT GCGAAAGCTG CTATGCAAT TGAATAGTT 720  
AGTATGCTC TGGGATTTT TCTCGAAGA TACCCCAAG GCGAGTTT TGTATATTA 780  
30 TATATGCTT TCTTCACT CTGGCTATG AATGTCTGT TGACTTAC ACCAGAAATG 840  
CAGCAGACC AACATATGG CAATGAGCC AAGCCAGTT GTTGGAGTCA TTGCTGTCT 900  
TTTTCTAGC TTGGATCTG CAGAGGCC AATGCTTAA AATTTCTGAG AACAGTGAC 960  
AAGATATTT TATCACTACA AGCTTTTAC TTTTATAGT ATTGACAG TATCTTACT 1020  
AATCTTCCA ATTTCTTTA AATGTATGA GTTTCTTAAA CAGCATATA TTTAACAGC 1080  
40 TCACTCTGC TTTATCTGAG TTATGTGCT CTATATATA TGTAGAGAA GATGTGGGG 1140  
TTTCTCACT CTGTACAGAC CACTGTATG TTAGGTGACA TTGATATGG GTTATATCA 1200  
GGAAGCTAA TTGTATTTAG TCGAATAAT AAAAAGTTTT TTTTATATA TTCACTCTGC 1260  
45 TTTTGGATT TCAATATATT AACTTTGCA AAGATTTTAC TTTGTACATG TTACAGGCTT 1320  
GATGTGTAA AATCTTTTAA TAAATATTA AATTAAGAA AATATGCA TTTTCTTTTC 1380  
50 TAAAAAAA AAAAAAAA CTGA 1405

(2) INFORMATION FOR SEQ ID NO: 184:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1596 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 184:

5 GTATCAGAT GCGCGGAGA ACTGTCTCT TTGAGGCGA CCGTAGGGG CCGAAGGGA 60  
AACTGGAGG CGAAGGTGAC GCGGAGCGA GCATTTGAG TCTGCTGGT AGACTGGTG 120  
10 CACACGACC ATGTGGCTG CAGGCTGGT GTGTCTCGG AACTACTCT CTAGGGTTTT 180  
CCACCAAGT TTACCAAGG CCGTCCGCT TGTGAAGAT TCCATCAGCA AGAATCAATG 240  
GCTTTTACA CTTAGCAGG AATATGCGAC CAAMCAGAA ATTGGATTC GCGGTGGAG 300  
15 AACTGCGCA GAATCAAG AGCAGCATT GGAACCATG ATCGAAMAA TATTTAAAT 360  
TATCAGATG GGAAGATGT TTGTGCTGG AGGCGCTGCT GTTGTCTTG GAGCATTTG 420  
20 CTACTATGC TTGGACTGT CTATGTAGT TGAAGTATT GAAAGGCTG TAATTTGCC 480  
TGTATATGC AAGATAGAA TTCAATTCAC CTATATATC TTACAGGGA GTATTTGGTT 540  
AAGAGTTTG TCTCCATAG CATACGAG AGCGCTGTT CTCAATGACT TCATGATAG 600  
25 AGGCTCTGG GTGACATG GTGTGACTT TCGACCATG GTTGGAGCTG GAATGCTGT 660  
AGCATATA CCAATATGACC AAGAGCCAG CCAAAAGCAT CTTGTCTGT TCGTACATTC 720  
TGTGTATG GTGTGAGTG TGGCTCTCT GACATATTA GGGGTCTTC TTCTCATAG 780  
30 AGCTGATG TACAGAGCT GCAATTTGG AGGCTCTTC ACTGTGCGA TGTGTGGCC 840  
CAGTGAAGG TTCTGAGCA TGGTTCACC CTTGGAGTG GCGTGGGTC TGTCTTTGT 900  
35 GTCTCATG GCAATATGT TTCTTCACC TACCAAGTG GCTGTGCGA CTCCTTACTC 960  
AGTGGCAAT TACGTTGAT TATTTCTTTT CAGCATGTC CTCTGTATG ATACCGAAA 1020  
AGTATATAG CTTGAGAG TATACCAAT GTATGAGTT CAAMATATG ATCCCATTA 1080  
40 CTGATGCTG ACTATCACA TGTATCATT AATATATTT ATCGAGTTG CAATATGCT 1140  
GCAACTGGA GCGAAGGAA AGAATGAG TCACTAGCT TCTGGCTTCT CTGCTACATC 1200  
45 AATATCTG TTTATGGGG CAGTATGCA TTAATATGT TGTACAGCA GCTTTGCTTG 1260  
AAGTTTAAA GATTAAGAAC ATGTATCAT ATTTAAATGT TCCGATATG TGAATGCTCA 1320  
GCTCTGCTT TTTTCTGGA GATTAATGC AATATATCT TCCAAATTA GCAACAGAT 1380  
50 TTTCAATCT CATGTGTAG TGAATTTAAA ATGTATTTGT GAATGTGAA ACTAAAGTTT 1440  
GTGTGATAG AATGTATATC TTTTCTTAC TTTTAAATTT AATAGTTCA CTGATATCT 1500  
55 AATATTAAC AAGCTGTGT TTGCATATTT TTGAGAGTG CAGATATATG TAATTAAGC 1560  
ATTCAATTA AATGTATTT AAGTTGATC TATATN 1596

60

(2) INFORMATION FOR SEQ ID NO: 185:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2293 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 185:

5 GCGCAGGCC CGTAGAGCA GGCAGAGCAG GACAGAGGCG ACTCCAGCA AACCGCGCTG 60  
ACCGTAGCG AGGAGTGTCT CCGTCGTGGC CTGAGGAGCC GCGAGGGTTA CACATGTTT 120  
TGGATGAGCT GTATATCATC TCGATTACCT GCGCTGATGT TAAITGATTT AGCATTTGAG 180  
CGAGGTTTAC AACTAGAGCC TTGTGGATMG AGACTTAAAG GTCTATTATAC AAGAAAGGTA 240  
ATCTGTAGCT CAGATGCTCC AACAGGGGAT GTTCTTCTTG ATGAGGCTCT GAGCATGTT 300  
AAGAAAGCTC AGCTTCGAG AACGTCGAG AACTGATMG AATTAATGAG TGGTGAACA 360  
TGGATTCAT TAAATTTGCA TTATCAGTTA AGAATGTATC GCGAGGATTT AGCTTAAAGC 420  
CTGGTGAAA AGGGTATTT GACAGAGAG AACAGAGCT TCTTACTTTT TGACATGACA 480  
ACAGATCCC TCACCAATTA CACATTTAG CAGGCTCTCA TCAGAGATG ACAGAGAGCC 540  
GTTCTTGACA AATGGTGAA TGACCTGAC GCGATGACA GCGCTTGTCT GCGCTTCATT 600  
TACTTGGCTC ATGCTCGGA GTCCTGGAG AATGCTTTTG CTCCTCTCTT GGCAGAGCAG 660  
TATGATTTGG CTACAGAG AGTCGGGAG CTTCCTGACT TAGAGCTGGA AGTGAGATGT 720  
CTGAGGCTCA ACACCAATGA GATTCTGTGG GCGGTGTGTG GCGCTTCAC CACTTAATCT 780  
TCTCGGGGTT GAGCATTTCT CTTTCTCTC AGTTAAACA GTATTTTTTC TTCTGTGTAC 840  
TTCTGGTTTT CTGPAATTTG TACTTTTCCA CACTATAATT GCTTCTGTT TTACAAATAG 900  
GTGGTGGCTT TTTTCTTTTT TGACGTGTA CAGATTTTG CTGGTAGAG AGGCTTTCTT 960  
CTTCTGTGTT TTAATAAAG TTATTCTGCC ATATTGGCAT TCCATTCCT GTTCCCATCC 1020  
TGACTGTATC CTGTTTTGG TTCTGTGTCT ACTTTGACTT TCAGATGAC TCCAGCTTCC 1080  
TCATAGCAC AGCTTTTGA TGACCTGAG TTGAGTTTCT CCATATGTGC ATGTACATCT 1140  
ACCATTTGCT CTACATTTCA GACAGATCT ACAGAAAGGC CTTCACTCA CCAAGGTAA 1200  
ATATCTGTAT CTATTAGCA ATTTTTTACA TAGACTTCAG TTGAGATGTA TACTTAGCA 1260  
AATTAATTTT AATTTGAA ACACAGTAA ATACTTAATA TAAATGTCC CTTCGATTTT 1320  
GCTTCCGAT TAAATCTATT GTATTATTAC ACTGTATATA ATTTTACTA TAAGGTGCA 1380  
ATGTGTTTAC AGAGCAGTT TGGATGGGC TGCATTCAT TTATGCTGTA TATAGTTTGA 1440  
ATTATATATA AATTAACCTT TCTTCTGGCC ACCGCTGCTC CCATCTTACT ATTTTGCAG 1500

ATCATATCAG TTGTACACT GGTGCGCTC GCTTGTCTCA ATCATGTTTA TTTCATGACA 1560  
AAATGAGCT CTGTGCGTG AGCGAGAG AGAAGATGTG TGTCTGATG GTTCTGGGAT 1620  
TTTTTGAGCT GTCCCATTTA TGTACTCTT TGCCTATGCA TCCCTTTTTT AGATTTTTTT 1680  
TAAATTTTAT CTTACTGTTT TTATATTTTC TATTTGGAG AGGCTTTTGA CCGATTAACA 1740  
TCTTGAGTTT CTTTTTCTGT CCAAGATTA ATTAATATCT GCTCTGAAT GTCATTTATC 1800  
TACTTACACA TTCTTGGGGA AAAAATATCA ATGTGAGTCC TACGATATGT TGCATGTAAA 1860  
TTGTGACAA GTATATGATA CAACCGAG GATTAGAT TTGTATACAG AAGCTCTAT 1920  
GTTTTATATT TTATATACCA ATTAGATTA TTACATATGT CAGACTATTA ACCTTTGCTT 1980  
TTTAAAGTTT AATTTTACTA TTCTTTTATC ACTTATATGT ATCATACCA TTGGTTCTAT 2040  
AATGTATATA CTATATGTTG AACAAATTA ATGTCAAAAT TTTTATTTAC CATATGCTAT 2100  
GTTTATAGTG GCGCTTTTCA GGTGTTAGAG ATTTTTTTTG TTGTGTTTAA CATTCTATCC 2160  
AAAGTACTA GATGCTGAT AACTCTAGAG TTGATTTTTTA AGGATTTCC TATATGTAT 2220  
ACTATCTTT TATCTGAGT AATAATATA CATGATCTT GAAATGCTC TAAATATATA 2280  
AAAAAAAA AAA 2293

(2) INFORMATION FOR SEQ ID NO: 186:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1212 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 186:

GGCAGAGGC GAGCGGCGC AGCTTACCT GCGAGGTGTG GTTCTACTC GTGCGCTTCC 60  
CGTGGGTTT GCGAGTTTA GCGACTATTG GCGCTGCGCC AGCGCGGCTG CAGAGCTGGG 120  
GCGGTGCTG CTGTGCGCG GTGATGCTAG CGCGCTTCTT GCGCTTCCAG CTGTTCGCGG 180  
GTGTAGGTGG GAGTACGGA CGTTTCGCG CCTGAGGTGT CCGCGAGGT GCGCGCATG 240  
GCGCGCAGCG GAGAGCATGG CTGAGCGCAT GGTCTGGGTG GACTTGGAGA TGACAGATTT 300  
GAGATTGAG AAGGACGAGA TTATTGAGAT GCGCTGTCTG ATAACTGACT CTGATCTCAA 360  
CAATTTGCTT GAGGTCTTA ACTGATTTAT AAAACAGCA GATGAGTTTC TGGACAGCAT 420  
GTGAGATTGG TGTAAAGAGC ATCACGGAA GTCTGCGCTT ACCAAGCGAG TGAAGAGAG 480  
TACATTTACA TTGCGAGCG CAGATATGTA ATTTCTGTTC TTGTAGCAC AGCAGACTCC 540



5 TCCAGGCTC TGTCACTTG CAGGAATTC AGTTCATGAA GATTAAGAT TTCTTGACAA 600  
ATACATGCC CAGTTCATCA AACTCTCA TTATGATTA ATGATGTGA GACTGTGTA 660  
AGAACTGTC AGAAGCTGCT ATCCAGAGA ATATGATTT GCACCAAGA AGCTGCTTC 720  
TCTTAGGCA CTGTATGCA TTATGAGAG CATCAAGAG CTTCAGTTT ACCGAAATTA 780  
CATCTTCAAG AAAAAATAG ATGAAGAA GAGGAATTT ATGAAATG GCGAAATGA 840  
GAGAGCTG ATGTATGCC ATTTATCAT CTGCCTAC ATCTTATCT GAGGCACT 900  
TCTGTGTT TTTTCTTC AGCTGATG CTTCAGAG CACTTCGCT TAATCTCAT 960  
CTCGAGTT ATTTACTAG CAGCAGAC AGGAATCT ATTTCTTC TAATCTCTG 1020  
TTTCAATAT GACAGAG CTCTTTGTA ATTAAGAT CATGTCCATC CTTCGTACA 1080  
TATATGCT TCTTTTAA CTTTCTTT TTTTATTA ATTAATATAG TAATTAAGC 1140  
TATTTCTAT GAATCGAA AAAAAAAA AAAAAAAA AAAAAAAA 1200  
AAAAAAA AN 1212

## (2) INFORMATION FOR SEQ ID NO: 187:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1605 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 187:

5 GCTTCGGAA GTTCTTTTG TCAAAATC CGGGCTTCT CTATTGTGT TCCGCGCAT 60  
CCCACTGTC CTCGACCTG GAGCTTACC TTCCGAGGC CAGATCTTG CCACTGCGC 120  
GGCGGGGCT AGCGGGGCT TCAAGAGAG GAGCTCTCA GGGATATGC AACTACAGCG 180  
GGCGGGGCG GCGGGGCGG AATGAGCT GCGCGGAT GGGAGGCTT CAGAGAAAC 240  
ATCAGGGCG GAGCTGAGC TTATTTGAC ATGAGAAC TCGATGATC CTCAGGCTT 300  
AGCTTCAGG ATTTGAGTGA GCTCATGAG CCGTCGCG AGGAGAGAT AGAGCTGAT 360  
GCACTGATG CAGCTGCTC TGAAGAGAG GATGAGAT TTCTGGCAT GAGGGCTTT 420  
AAGGAGAC TGAAGCGCA GGTGCAAT CAGATGAG AGCTGCGAA AGACAGGCC 480  
TCAAGGGCT TCACTTTGA CCGCAATC GACATCTCA GACCTACTI TGAATGAG 540  
CTCTCTAGG TCGAGAGG GCTCTGAG TCGATATCC CTATCAAGT GTTCAGTTT 600  
CCCAAGAAA TTGAGGTGA ACTATGGA CCTCATATC TGGTCTTAC TGTGTTGCT 660  
ATCTACTTC ATGGATGAA GAGCTCTAC ACTATATTC GCGAGGGAC CCGATGCGC 720

5 AAGGCAATG GCACTGCTT CCGCTACTG CCGGAGTCT CATCTTCAAT TTACTTCTT 780  
GCTTACTGT GCAAGCGCA GATCAACATG CTCGAGTGT TCGACTGCT GGCATATGC 840  
CTCTTTGGC ATGCAATGT CTTTCTATC ACTATTAAT TCACTCTCA GCGCTCTTC 900  
TACCTCTCT GCTTGTGTT GGTGGAATG TCCAGACTG GATATGATG AGTATTTGTG 960  
TCTCGAGCG TGGGCGCCAC AAGGCGCTG CTCTCTGAT GCACTCTGC TCGCTTACAC 1020  
ATGCTCTTC TCTTATGCT GCAATTTGCG TACACAAAG TGTATAGGG GATCTTGAC 1080  
ACATGGAAG GCGCCAAAT CCGCGCATC CAGAGGCTC CAGAGACAT CCGTGCATG 1140  
CTCTCTGCT CTCGCTTCC CACCACTTC CTCAGCGCA CAGCAGAGC TTTTCCGCTG 1200  
ACCTGCACT CAGCTGACC CCACTGAAA TTCTTGCCA GTCTCTTTC CCGAGCTGC 1260  
AGAGAGAG AAGACTATTA AAGGAGCTC CTATGACAT GTTCTGTAGA TCGGCTTTC 1320  
AGCTGCACT GAGCTGATG TCGATAGTA CTTCTTAT CATTGTGCG ACTTCTGAA 1380  
GCAACAGCG CAGAGCTCC TCGCAGAC TCGAGGCTC TCGAGCAAT CAGAGAAATG 1440  
GCTCAGCTC TTGAGAGC CTTTCCACC TACCGCTTC TTCTCTTTA TCTCTCCAC 1500  
ATTCTTTC TAATATAGA CTGATTAAT AAATGTGTA TTGAGCTG GAAAAAAA 1560  
AAAAAAA AAAAAAAA AAAAAAAA AAAAAAAC TCGAG 1605

## (2) INFORMATION FOR SEQ ID NO: 188:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1516 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 188:

5 ATTCGCGAT AGGGGTCAC GTGTGCTG GCGCGGGA ATGGCGCTT CAGCAGAG 60  
CGGACTTCA GCGCGCGAG GCACACCA GAAAGCAT TTATGACTTCT ACATGAGAT 120  
GAACAAATA GAGAGAGAG ACTGCTTCT ACTTGGAA ATTCAGATTG AAGACTGAC 180  
CGTCTGCTT TCTCTTACT TCAATTGAA CCAATTGAG GTTCTTACA TAGATCTGA 240  
AGTTACAGT GAGAAATTA AAGAGGTT TCGCATTTA TCTATCTTG TCAATCTGA 300  
CAAAATCAA GATGATCTG ACAGACCA AAGGCTTTT GAGCTGTG ACAAGCTTA 360  
CAAGTTGTA CTGATCAG AGCAAGAA GAGGCGCTG GATTTATTC AGCGAGAAA 420  
AGATATGTC GACATCTG TGAAGAGCG AAAAAACA TTAAGAGAG AAGAGAAAC 480

TACATTTGTA GAGGAGGATG ATTCGAGCT GTTCAGAGAA CCGTGTATATA AACAGAGAT 540  
GAATCTTTT GCAGAGCTG AAATTAAAG GAAGAGAGA GAAGCCAAG AGATGATCA 600  
5 AAGGAACCA CAAGGAGG AGAGATTGA AGCTCAGAA AAGGCCAAG GGGAGAGAG 660  
GTGCGAAA AACTTTTAGG AAATCGAGA TCGTGTGTG GACAGCTGC GAATTTCA 720  
AGCAATAG AGGGGAGA AAGAGAGAA AAATCGACC TTCTGTAGAC CACCGAAGT 780  
AAAAATGAG CAAGTGAAT GACCGCCAA GGTCCAGGC ACAGAACTT TCCCTGCTA 840  
TCCTCCTTC TCTTTCGAG GACTATCTT TCTCTCCAC TTCCACCCCA ACHAGAGTA 900  
15 GTATTTGCTT TTATGCTAT TTGTTTCA ATACGATTA ATATGATCA GATTAATCT 960  
TTGTACATT GAATGAGG CCGTGTATTA AAAAGAGCC TTTCCTCTC CCGCCCTTA 1020  
GAACAGCAG TATTAGAGG TCCACATAT GGTCTGCTT TCTCTTCCA CAGCTGTAA 1080  
CTCAGTCTT TCTACTTAC TGAATTGTA TGGTTAGAAA CTTCTGTGAT AGTTTGTGA 1140  
AATCATCCA TTAAACATAC TCGTTAAAC AGTGTCTG TGACTTCAGA GACAGCTG 1200  
25 GAAGGECAC CTTAGAGAG CCGTCTCTT CAGTGTCTG CTTCTGGTG TCTCTCTTC 1260  
GAAGGECAG ATAGAGAGG GAACTATGT GAGCAGCAG AGCAGCATCT GATCCCTGT 1320  
GCTGTTTGC ATGTGCCCC TGTCTACTA CCAATCATG TGGCATGAG CCCAGCCAC 1380  
CCAAACCTT CAGTTTCCA AGAGCTAGC GTCTCTCAC CAGTACCATG TCTTAGCTG 1440  
TCTCATTTG TTAGTGTAA TATCTTTAT GTATATAAA TTTTATTAC CAAAAAATA 1500  
35 AAAAAAAAA ACTGCA 1516

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 681 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 189:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 189:

60 GCTCCCATCT TCTGTGCTGT CCGTATACA CCGTGTCCC TCCAGGAGG GCGTACAGC  
CATCTCTCTC CTGTAGGCT CTAGATCCC TCCACTTTT GCGCTCTAG TTATCTGGG 120  
CAGGGGACCA TTGAGCATC CTCCCTCTT CNGAGCTCA GGTCTGAG TATTAGCCCT 180  
GGGCCCCAGA TCCGTGTAA CACTCTCTG GAGAGACTC TCAAAAGTGA CTGTATTTT 240  
GAGTTTACCA GGAATATCT CCGACACTG AAGCACTGC AAACCCWAG ATCCCAAGT 300  
60 CTTTGGCTC TGTGGACTG TCTTCCGAG ATCTTCTG TTGCATATG GGAAGCTTAA 360

GAGGAAAAG AAGAGGCTT CCGTCCGAG CCAATCCAG GATTCATGC CCACTGCCC 420  
TCTCTCTCC TCGCTGATAT GTGGGCCCC GCTCCGCTTC AGGTTGTGCT GTCTCTGAC 480  
TATGTTTACA TCCCGGAGG GTTCTGCTT CCGTCCGAG CAGGTCAAGG TGTGTTCCAG 540  
CAGCTTCTG TCGGCTGCTG AATGTGTCA CCACTGCCC CTTTCCCCC GGGGGGCTCA 600  
TGGTCTCTC CTGATGCTG CCGCTGAT YTTTTTTTT GAAAGACCTT TTAAATTTA 660  
AAAAAAAA AAAAAACTG A 681

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1014 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 190:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 190:

60 GCGTCAAGC AGCATATGA TAAATTTCT GAAATTCMA ATTCAATGTT TCTACAGCA  
GTATGTCTAC AAACCATTC AGCAGACCA TCAACACGA GTCTGCCAT TTTTGTATC 120  
GCTGATGAT TCGTGGCCA AGTGGCTT TTAGAGCTT TAAAGGCCA CAAAAGGCT 180  
ATCTGTAAG CACAGTCAA CAGATTTGT TATATGCA AGCCATTCG CCGTATGAT 240  
TATGTGCTAC ACTCTGAC ACCTCAGAG TCCAGAAAG CGAGAGACGA GTTGTACCA 300  
CTGTAGCAT AGCTATGTT GCAGAAAT GTTCAGCTT TACAATATC CCGCTTTAA 360  
TGAATGATA CAGATTTCT GAACTGATG TTCAAAATG AGTGTTAAA TCGCTTTCT 420  
TCTTGTTTA ATATATGTT GAATGGGAA AAGACTAGT TTATGCCCTA ACAGCTTAC 480  
TTGAGATGC TTTATGATAT AGAGACTTG TACACAGACA GAGCGTATG CCAATGTGAC 540  
AGCAATATC ACTTGGGTT TATGATTTG GTTGTGAGA TTGCTGAT CACTTTTGA 600  
ACTATGATG GCGATGTT TTTGAGCAT CCGCTCATG AATTCAGCA GTTATGGAG 660  
CCCTAGAGG CCGTAGAGT CATTATGAC CATGTAGAT GTTCAATAT TTTTACAG 720  
GTCTTTTCA CCGACCCCG AAATCTGAG ATGTATATG GAATTTTAC AACTCATCT 780  
AAATTTGTC CGAGAGGCT CTCATACAC ATTACCGAG AATCTACCA CAGATATAG 840  
TACAGCTATA TCGTTATGA ACTGACTAT ACTTATAT TTTATTTT ATTTATGCT 900  
TATGCAACG TACTTCAC CTTAACTTG CTTTGTATG GTGATGTAA CTTTAAACA 960  
TTGCAATCA GTTACAGCT GTTCAATAG GGAAGACTA GGAATTCAT AGC 1014

## (2) INFORMATION FOR SEQ ID NO: 191:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2779 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 191:

5 TCCGACGAG CTGTGTTCG ATGGTCAATC TCTGTGTGGT AGCTGTTCCT GACAGGGGAG 60  
AGTTAAGCTC CCGTCTTCCA CCGTCCCGCC TGGCCAGGTG GCTTGAGGCT GACCCAGAGA 120  
CCGAGACTTC CTTCCTGAG CTATGTGCTC AGAGCTGGGG AGGAGGGTTC CCGCCGTTCCT 180  
CTGCTGTGAG CCGCGGAGCC CCGTCCCGCC TTCACTTCTCT CCGCGAGCCC CTGCTACTGA 240  
GAGCTTCGCG GATTCGACGA CCGCGAGCC CTGTGCTTCA GCTTCGCGGG CTTCAGTTCAG 300  
GCCAGACCG AGCGGCAATG GAGAGAGAC AGGAGCCCTTG ACATCTTCAT CTGCAAGGAG 360  
GTCTGTGCTG GACCGAGCAG CTCTCTCTC CTAGAGTAC CTACCCCTCC AGCTTCTCAG 420  
TTTTCAGGTT GAGAGATTA GATGAGGCC AAGAGATGCG CTCTGAGCGG GACAGAGGAA 480  
AGCTGAGTTT TGGAGCGGG CTGCTTCCA TGGATTCACA GTTTCAGGCC GAGAGCGGGA 540  
AATTCGCCC TCGATTAAGA GTCACTTCA ACTACCGAAA GGGAGAGGCT CCGAGTTCAG 600  
CGGATCGAAA CCGATTTCAG CGAGTTCGCG TCTTCAATGC GTTCTCCCGG GTGTCTCCCG 660  
AGGATCTGCG TGGACTTCCA GATTACTTGA GAGAGACAG CAGTAACTTC ACCGACTCGG 720  
AATACAGAGA GGGCTCCACA GTTAAGAGT CCGTGTGAAA GCGTGTGCTG ACCCTTAAGG 780  
AGGGGTCAA TGGCTCGATT CTGCACTGCG TCGAGTCCA CCGGGACTCT GCGAAATCTC 840  
AGCGCCCTGG AATGCGCCAG TCGACAGATG ACTATTACCG AGGCGCAGCG GCTCTGCGAA 900  
TGGCAATTGA GAGAGGAGM CTGAGTGTG TGAAGCTCTCT GTTGAGAAAT GGGGCGCAATG 960  
TGCATGCGCG GGTCTCGGCG GTTCTTCCA GAGGGGCCAA GCGACTTCTT TTTATTTCGG 1020  
TGAAGTACCC CTCTTTTTCG CCGCTTCGAC CAGGAGTGG GATGTGTTAA CTTACTTCTT 1080  
GGAGAGCCA CACCGCGCG CCGACTTCCA GCGCTGACT CCGAGGGGCA CAGAGTCTCG 1140  
CATGCCCTAG TCAATATCTC GAGAACTCA GCTGAGAAC TTAGCTTGGT GACCGAGATG 1200  
TATGATGGCG TCTTCGACG TGGGGCGCG CTCTGCTTCA CCGTGCAGCT TGAAGCAATC 1260  
CGCAACTCC AGGATCTCAC GCTCTGAG CTGGCGGCCA AGGAGGGGCA GATCGAGATT 1320  
TTGAGGACA TCTTCGAGCG GAGATTTCGA GACTGAGCC ACTTTTCGCG AAGTTTCAC 1380  
GAGTGTGCT ATGGGCTCT CCGGATTCG CTGTATGACC TGGCTTCTGT GAGACAGCTGT 1440

1500 GAGGAGAACT CAGTCTGTGA GATCAATTGC TTTCATTTCA AGAGCCGCGA CCGACACCCA 1500  
1560 ATGCTGTTTT TCGAGGCTCT GAGCAAACTG CTGACGGCGA AATGGAACTT GCTCATTCGCC 1560  
1620 AAGTCTTCTT TAAACTTCTT GTTAAATCTG ATCTACATGT TCACTTTCAC CCGTGTTCGC 1620  
1680 TACATCAGCG CTACCTGTAA GAGCAGGCC GCGGCTGACC TGAAGGCGGA GGTTCGAAAC 1680  
1740 TCAATGCTGC TGAAGGCGCA CATCTTATC CTGCTAGGGG GGAATCTACT CCGTGTGGGC 1740  
1800 CAGCTGTGCT ACTTCTGGCG GCGGCAATG TTCAATCTGA TCTGTTCAT AGACAGCTAC 1800  
1860 TTTCGAAATCC TCTTCTCTTT CAGGCCCTG CTACAGTGG TGTCCCAAGT CCGTGTTCCT 1860  
1920 CTGCGCAATG AATGTACTCT GCGGCTGCTT GTGTCTGCG TGTGTCTGGG CTGCTGTAC 1920  
1980 CTGCTTAACT ATACAGTGG CTTCAGGAC ACAGGCAATCT ACAATGTCT GATTCAGAG 1980  
2040 CCGTGTGAG CCGTACGCG GANNITGGG CCGGAAAGCT CATTACAGGC CGAATGCGAC 2040  
2100 AGATGAGTGG CAGGCAATGG AGGAGACGA KGAAGAGGCC AACGGGCGCC AGTACAGGGG 2100  
2160 TATCTGTGAA CCGTCTCTTG AGCTTTTCAA ATTCAGCAATC GGAATGGGCG AGCTGTGCTT 2160  
2220 CAGAGAGAG CTGCACTTCC GCGCAATGCT GCTGTGCTGT CTGCTGCGCT ACCTGTGCT 2220  
2280 CAGTAACTC CTGCTGTCTA ACATGCTCAT CCGCTCTCAT AGCGAGACG TCACACATGT 2280  
2340 CCGCACTGAC AGCTGTGAGA TCTGTAAAGT GCAAGAAAGC ATCTCTGTGC TGGAGATGGA 2340  
2400 GATGTGTAAT TGGTGTGTA GAGAGAGCA GCGGCGAGCT GTGATCTGTA CCGTGTGCGAC 2400  
2460 TAAAGCAAT GCGAGCGGCG AATGAGCGCTG GTGCTTGAGG GTGAGAGAG TGAATCTGCG 2460  
2520 TTGATGGGAG CAGAGGCTGC CTAGGCTGTG TGAAGAGCGG TGAAGGGGAG GTGTCTCTCG 2520  
2580 AACTCTGAG AAGCCCTGTCC TGGCTTTCGC TTCCAGAGAG GATGAGATG GTGCTCTCTA 2580  
2640 GGAAGCAAT CTGCGCGTCC AGCTCTTCCA GTCCAACTGA TGGCGCAAGT GCGAGAGGAG 2640  
2700 GCGAGAGAC AGAGCAGAGG ATCTTTTCAA CCAATCTCTC TGGCTCTGGG GTCCCAATGA 2700  
2760 ATCTGTGCG CAAATATATA TTTTCACTTA CTCANAAAAA AAAAAAAAAA AAAAAAAAAA 2760  
2779 AAAAAAAAAA AAAAAAGGC 2779

## (2) INFORMATION FOR SEQ ID NO: 192:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1923 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 192:

ACCCTCTCG CTCGCTCCG CTCGCTCCG GCTGCTCCCT CACATGATC GCTGCTCCG 60  
TGGCTCTCGA GCTGCTCCG TGGATCTCTG CCTGCTCTCT ACTGAGGCC ATGCTCTTGG 120  
ACATCATGC GCTGCTCCG GCTGCTCTCT TGGATCTAG GATCCTGAG GATGCTCTCT 180  
GCTGCTCTG GATGCTCTCT CAGAGGCGG GCTGCTCCG GCTGCTCCG GATGCTCTCT 240  
AGAGCTCTAT GATGCTCTCT TGGATCTAG GATGCTCTCT CAGAGGCGG GATGCTCTCT 300  
TATGCTCTCT GATGCTCTCT ATGCTCTCT TCTGCTCTCT CAGAGGCGG GATGCTCTCT 360  
TCTGCTCTG AGTATGATG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 420  
TCTGCTCTG AGTATGATG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 480  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 540  
GCTGCTCTCT CAGAGGCGG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 600  
CAGAGGCGG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT GATGCTCTCT 660  
TGGATCTAG TGGATCTAG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 720  
AGTATGATG TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 780  
AGTATGATG TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 840  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 900  
TGGATCTAG TGGATCTAG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 960  
TGGATCTAG TGGATCTAG GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1020  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1080  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1140  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1200  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1260  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1320  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1380  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1440  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1500  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1560  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1620  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1680  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1740  
CTTACATCTA TACTGCTCT GATGCTCTCT CAGAGGCGG GCTGCTCCG GATGCTCTCT 1800

CAGATCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1860  
CCATATATCC AGTATCTCTA AAAAAAATA AAAAAAATA GAGGCTCTCT GAGGCTCTCT 1920  
5 ATT 1923

10 (2) INFORMATION FOR SEQ ID NO: 193:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1.1) SEQUENCE DESCRIPTION: SEQ ID NO: 193:

AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 60  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 120  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 180  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 240  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 300  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 360  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 420  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 480  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 540  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 600  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 660  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 720  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 780  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 840  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 900  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 960  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1020  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1080  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1140  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1200  
AGCTCTAGGG GATCCTCTCTA TTTTCTAGT TGTGAGAAA GCTATCTCTA GTTCTCTCTA TGTGATCTCT 1260

5 AAAAAAAAAA TTGAGCAAC GTTGGGAC ATGGTTTTTG AAGGTTTTAG TTCTGCTTTC 1320  
TGCCTCTCTT CCAAGACCC AACCTCCAC CCTCATACA TGAGCCAGTG ATATATCTTG 1380  
TTGAGGAGA AGATCAATTA GATTTGTTTT GATTTCTTTA GAATGAGGG CAAATTTCCA 1440  
CAAGTCCCTT GCTGTGATG AGTTGCTTG CAGGGCCCG AGTAGAGCA CTGGGGTGG 1500  
10 GCGGAAATG GGGTTACTG ATGTAAAGGA TTCTTTTGG TTGTGTTGAG ATCCATGCA 1560  
GTATGATTT CTGTGATCC CAGTTGGTT CAGGAATTT TGTGTGATG GCTTAATCC 1620  
15 AGTTTTCAT CTTCAGAGC TGGGCTGAA GTTAACTCA GTAGCTGAC CTGTCTGACC 1680  
CGTCAAGTT CTTCAGCTT CAGACTTT TCTCTTTGTC GGGTGGGG TGCGAACTCA 1740  
CGTGGAGC GTTGGCTGAG AAATGTAG GATTTGAA TACATATCC ATGGGACTTT 1800  
20 CTTTCTCTT CTTCTCTCT CTTTCTGTC TCCTAACCT TTGCGCGAAT GGGCAGCAC 1860  
CACTGAGTT TGTGGGGGC CAGTGGGCT GCGAGTTCC TGTACTACTG CTTGTACTTT 1920  
25 TTCAATTTGG CTACCCCTG ATTTTCTCAT AGCAAGTTTG GTCAAGTCA ATTGAATATT 1980  
GTATGTAGC CAGTGGACC CAGCAATTC TGGACCCCG CAGTTGGAG GAGGAATAG 2040  
TCAAGCTTC CAGTTGGCT GAGAGCAAT GACTGTTAC CTGCGGCCA TCACCTTGA 2100  
30 GCGTTCTCT GCGTTGAGT AGAAATGTG GGAATCGGG CAGAGAGCC TCAATACGA 2160  
TGGAAACTA TTGTGACAA GTTCTTCAG AGCAATTTCT TAATGAGATA TTGTATTTA 2220  
TTTCAGACC ATTAATTTTG TACTTTTGA AAAAAAAAAA AAAAAAAAAA 2280  
35 AAAAAAAAAA AAAAAAAAAA CAGGGGGG CCGTACCCAA TTGCGGTAT ATGATGTTA 2340  
ACATTC 2346

(2) INFORMATION FOR SEQ ID NO: 194:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 3054 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 194:

55 TATCTGACC ACCCTTATT CTACATACA TAGGCGAC TGAAATATCC TAAACCCCTA 60  
AGCTGAAGT GCGCTGTGAG AGAGCAACT GAATATAGC AGGGCTGGGC TGTGTTTCC 120  
TGGTATAGG CTCACCTTT CCGCAATTC TTCTCTGGA GCTTTGAGC CAGGTGCTA 180  
60 AAGGAATG GTAGGAGC TCTTATCT ATTTCTTAA AGCAATATCT TGAAATCA 240

5 TTCAAGCT GATGCGCTAT AACCGCTGCC TGAATTTCTT CTATTAGGC TATTAAGAT 300  
AGCAAGACT TTACATATT CAGATGTGTT TCAATGCTT CTACCGCTT CTATGCGCC 360  
CTCCATTTAT TTGACTAAG CATAACAG TGCACTAGC ATTATACAA GAGTATAGA 420  
AATACAGTGC TTATGCGCTC TACATTTACT GCGTTCAATA TCAGGCTGCC CTGAGGAAG 480  
GATGCGACC TGAAGGCTTC CTTATGTCTT CACCAAGAG AGCTGCTTCA TGAAGGTAT 540  
CTTTTCCC TATGCTGTTT TTGCGCTGCC GCGTCTTAT GTTACGTGGG TACCGAGCT 600  
GGTTCTTGG CTAGGTATG GGGACCAAT TCAATACCT CATTCAATTT CTACATAGT 660  
15 AAATACGCT ACCATGTTA GTGGAGAG CTGGTTTTTC CTAGTATACC CACTGCAATC 720  
TACTCTTACC TGGTCAACC GCTGTTTCCA GTTATGGAG CTGCTAAGTG TGGATTAAC 780  
TGTATAGGGA GAGGGAATA CAGAGAGGC CTCTGTGTTT CCGTGGCTCA GCGAGCTGCC 840  
20 CAGAGCCT AAACCAATTA AACAAGATA CTAGTCAAT TTTTATCTG GCTTCTCTTC 900  
ATTCGACTG CACTGTGTC TGGTTGGCT GACTGGGAG ACCCATTAC TACAGAGTCT 960  
25 GACGAGGA CTGAGACTG TCACTTTCTA GCTCGAAT TACTGTGTA ATTAATCTTC 1020  
AGAACTCTA CCAATAGTG AAATGCGAC ATTTGCTTT ATTAATCTCA CCAATGTGG 1080  
GAATACTGG CTTTTTTCCA GCGTTTCCA GGGCAATAA CTCAGCCCT TCGATAGCA 1140  
30 GTCCATCAG CCAATATTT TTTAAGAA ACTTGCAT TGTTTTCTT TTTACATTA 1200  
CTCTCTCTT GCGCAAAAT TATTAATCTT AAGTGTAAA AAAGTCTTA ACACAGCTT 1260  
35 CTCTCTTAA AAATATGTA TTAATATCT GTATTTTAA ATTCTGCTC TGAATAATGA 1320  
CTGTGCAAT CTGCACTAC TGAATTTGG GCGTTTCCA TTGCTCTCA TGTCTTTTAT 1380  
CAATGAGGC CAGTGGACAG AGGAGAGG GAGAGAGG GTCCCAACA CTTGTGTTG 1440  
40 TTCTGACTG ATCTGACCA AGAAGAGTA ACATGAGC GCTGCGTCC ATGCAAACT 1500  
CTCGAACA CTTATCTCT TCGAAGTG GCTTTTCAG GCTCTTTACT GCGAGCAAT 1560  
45 TAGAGCCCT CCGAGCCCT TCGTTTTC TTCTTTACT CTTTGGCTT CAAGATTTT 1620  
TGGAAAGA ACAATGCT TTAACATAT TTTCAATTC TAAATTTCA GGGATATG 1680  
1740 AAAAAAGG CAGTGGCTT AAGGCTGTG TAAAGTGA GAGAGGAA ATCTTAAGT 1800  
TACAGATA AAACGATC CCAACAA AAGCAAT AGAATGCTC TTCCATTTG 1860  
CGACTTTCT TGTCTATAC AGCTATAC CTGGAGAG TAAATTTCA TTAACCAAG 1920  
55 AAGTGGCT ACTGACTC TGAAGCTG ACTACTAG CACTCCAT CACTCTACA 1980  
GATCGAAG AGTTCGAG AGTCCAGT CTTTAACTG AGCTAGCA ATTAACCTG 1980  
GCAATGAG CAGAGAAAT GAGAGAAAT CCACTGTGA GGTGACGC AAGATGAAA 2040

5 GAGGAGGAG GAAAGAGGTA TGAAGGCGAG AAGAGGATTC ATTATGATGG GTCTGAAAGG 2100  
AAGAGCTTTT GCTATCCGAC ATGTACTGCT ACTACTCTGTA AGCATTTTAT GTCCGAGAT 2160  
GGAAAGAAAA ATGAGCTATT GTTATATATA TATATGCTTT TCCCTGGAGT CAGTTTATTT 2220  
AAGAGGTAAA CTCTTATTTT TTACTGTGTT ATTTCTAAAA GAGAGGGGAG CTGAGGCGAT 2280  
TCCCTGTAGG AGTAAGATTA AAGAGTAGG AAGAGATTC AAGCTCTTAT AGAGTCACAG 2340  
CTTTCCGAGG TATTAAGCTT AATATTAAGA AGTACATATA GAGAGGTGG AATATGATCT 2400  
AGTCTCTGAT AGTACCTCAC AGAGCAAGTG ATTTATTAAT TTGAATATCA AACTACTTTC 2460  
TTAATATCAC TTGTGCTCC ATTTTTCGA GAGAGGAAA TATGTCCGCC CTACTACTTC 2520  
TTCTTCTAAA ATTAAATTC CAGCTTCCA AGATCATTTCT AAGATTAATT TTGCACAGAC 2580  
ATCTCTCAC CCACTGCTT GTCTGAGCT CAGCCAGGT CAGCAACAA CTTCGTGTG 2640  
AAGCAACTG CTTTACCTT CTGGGGAGG GAGATTAGT AGACTAGAG ACCAGAGTG 2700  
ATGGGAAG GTTGAAGT TCAATATTT GGCCTGTGAG AGCTTGATTA GAAGCCAGA 2760  
CAGTGGCAG AAGGAGAC TTGGCCGAG AAAAACTGT GGTGTGTGCT ATTCTGTTC 2820  
CAGAAATAG GTTGACAGA AGCTTGTGG GTGATGGAG GATTTGGAG CTGATATGT 2880  
TGTATATTC GAGCTGTAAA TTTTGTGAT GTAAAGACGA ATATCTGTGA AACTTAATGT 2940  
CTGTATATAT ATGAGGTTT ACACAGTAA ATATATCAAT AAGAGTCAA AAAAAAATA 3000  
AAAAACTGG AGGGGGGGCC GGTATCCGA TTATCCAAAT AAGATATGTA TTAC 3054

(2) INFORMATION FOR SEQ ID NO: 195:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 907 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 195:  
GGCAGAGCT GTGGCCGAAA CTTTTCCTGC TCTGTGCTGC CAGCTACTGG CTGGCCGCGG 60  
CCCTGGCTTG GGCCTGCACC AGCCTGGGAG CGGGCTTCCA CAGCAGCCCC CTTCGAGCA 120  
CGCTGCCAC ACCCGGCACC TTCTGGGGA AGTGTCTGCG GTTCCGCGGG ACCATATGA 180  
CGAAGGCTTT TGTGCTCACC TACAGCTGG GTAGCAGGG TTCCAGCAGC GTTTGATTC 240  
TCTTGTCTCC TGTGTGAGCA CAGCGGCGCT TTCTGTCTCC GAGTTGGGAC TTGTGATGG 300  
TGTGGTGTCT GTGTGTGCT CAGTGTGCG CTCTCTCCCT GGTGGAGCT TGTGTGCTAA 360  
GCACTGGAAA CTCTGTCTC TTGAGGTGG GTGTGTGCTT TCCGCTCTGG GGGCTTAGCC 420

5 TGTCAAGTCT CTGTGTCTT CCAGCTTGGC CAGCCTGGGG GCGAGCATGG ACCTGTGCAC 480  
AATCTTGAGA GGTTCAGCCT TGTGAGCCT ATGTCTGCG CAGTCTCTTG GARGCCTGCT 540  
CAGCAGATC ACCTTCACTG GGAATGATGC GCTGCAGCCA GCTGGCCCC AGGGCTTTTC 600  
AGGCGACACA CTACAGCTT CTGGCCAGGC TGGAGCTGCT GGGGAGCTG CTCTGAGCCA 660  
CTTGTGSSCG AGGGCTGGC TGAATGGTTG GGGGAGCATC CTGTCTCTTT GTTCTGTCTC 720  
ATCTCTCTG CTTTTCCTT TCTGTACCTG GAGCTTAGCAC CAGCAGCTT TCTCTGAGCT 780  
GAGTGGCTGG AGTGTCTCAT AAGCCATAT GTTGTCTTGG CCGAAGAAA AAAAAAATA 840  
AAAAAATA AAAAAACTG GAGGGGGGGC CCGTATCCCA ATTCGCCGA TATGATCTTA 900  
AAGATTC 907

(2) INFORMATION FOR SEQ ID NO: 196:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1290 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 196:  
GCGACAGGA GCGACAGGA GTTGGCAGG GGAAGAGCA GCTTATTTCA CTACACGCC 60  
CTCTGTGTGT CAGCAGCCT CTGTGTGTGG TGGAGAGGCT CTGAATCAG AGGGCCCCAG 120  
KAGGGCAGGA AAGATGGGG CAAAGCTGG CCGTCCGGCG GGTCCGGGC AGCTTTGCHA 180  
TCTGAGCCA CGCTCTCTCC AGGCAATGCT CTTTGAATTT GGAATATCA ACCGAGGCC 240  
TTTACACCG CCGTCCAGCA TGTATATGAC TTGAATCTAC TCTAATAGAA TATTTAATCC 300  
AAGCTCACT AGATTGAGC TGAATCCAC GACTAACCTT GAATGGGGG TTTTCCAGCC 360  
TTTACGCGGA TGGCAGCG GTTCCCTGGG GGTCTGTGCA CGGGCTTAT CTTTCTCTGT 420  
TCCAGACTTT GCGCTCCGAC CTCTCTCGCC CCGATGGGCT GAGCCCGTCC GTTCTCTGT 480  
CTGTCTATAC GTTGTAGTCC AGCTAAGAG ACAAAAGCA ACCGTTGGGC CAGCTTGGGA 540  
AGGTGGCTGG AAGAGCTCC GAGTGTCTGG AGTGTAGCC CTTGGATGG CATTTCCATTG 600  
TGTGCTTAT TCCGTGAGAA TCTGTATAGC GCTTGGCTTAT AAGAAATTA GCTCTCTCAT 660  
GCTGTATTTA AAGGACTTTT AAAAAAATA AAAAAAATA AAAAAAATA GGGGGGGGCC 720  
GGTACCAAT TCGCCCAATA GTGATGTGTA TTACATTTCA CTGGCGCTC GTTTTACAAA 780  
CGTCTGCAC TGGGAAAAAC CTGGGTTTAA CCAACTTTAA TGGCTTTGCA GCAATGCC 840

5 CTTTGGCAG CTGGGTTAA TAGCGAAA NCCCGCAC CCATGCCC TTCCACAG 900  
TTTGGCAGC CTGATGCG GAATGGCA ATTGTAGCG TTTAATATT TGTATAAAT 960  
TCCGCTTAA AATTGTTGT TAACTGCT CATTTTITT MACCATATA GSCCAATTC 1020  
CGCAAAATC CCATTATTA TTCCAAAA ATAAACCAA AANGGTTTG AATTGTTT 1080  
TTCCCAATT TTGGAAACA AATTCCTCT TTATAAAA OTTGAAACC CAACTCTC 1140  
AAGGGGAA AAGCTTTT TTGGGGGA AAGGGGCCC CTACTTTTA ACATCCCC 1200  
CCAACTATT TTTTGGGG GTCCCAAG GTCCCTTAA AACTTTTTT CGAAACCA 1260  
AAGGAAACC CCATTATA AATTGCTT 1290

(2) INFORMATION FOR SEQ ID NO: 197:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1020 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 197:

30 GGTGCTGCT GATGCTGCT TAGGTGCT TTACCAAGA TTATGTTAC AATGATGA 60  
GACTCTATG GAGAAATAT TGAAGNAT TAAAGAGAC CTCATATAT GAGAAATG 120  
SCCTTGAGG AATTGATATG AGCTTTACA GTATCTATT TTCACTCAA GCAATGCT 180  
TTCTACACA ACTCTAATC ATAAAGGCT CTATGCAAT CTATGAAGA GTACAGAC 240  
ATGCTTAGTG TGTATCTC TCTTTGAGC ACTGTATTT CTACAGAAA TTTCCAGAC 300  
ATTATGAGG TAGAAAAAA TCGACGAG CTGTTAAGA TCTTGATCC CATATATAG 360  
TATGTATAG TGAATCTGT AATCAATCA CTTTTCCTT TTTATCTCT TACCAAAA 420  
TTGTTAAT TTGATCCA AATGTTTAA ATCTTGAT ATTTTATA AATCTTTTC 480  
TCTCATCAT TCCCTTTT STGGTTTAA ATAGACTTAC TTGCTTTTG AGATGAT 540  
ACTCTTCT ATCTACAAA TATGTATAT GGTATTTTC ATACAGATG TCAATTTGA 600  
ACCAAGATT GGTATGTT TTTAAGAAA AATCTGCT TCAATCTCT GAATTTCTC 660  
TTGAAATTT TCTTTTACA CGTTAGGCC AACTGAGTA CCTGATGCT GTTGATTT 720  
TTCAATGAT CTACACTCT ATTTTAGCA CTGACCTTT TATATTTCT CTATTTTAA 780  
AGTTATTTG TCTTAACTA TTTAATAT ATACTTTTA TCTGTTTCTG AATGGGACT 840  
GAATTTTTG GATATGATA TTGATTTGA AATTTTGG AATTTTTCT ACTGAAAT 900  
TTAAGAACT AATGAAAT TCTATATG ACTGAAATG AGCTTTGTA CAGTGACAC 960

TTCTAATAA TATGATGCT TCCCTAAN GAGGCGGAC ATGTCCACT TTCCACAG 1020

(2) INFORMATION FOR SEQ ID NO: 198:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 524 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 198:

15 AATCCCGAA GCTAGGTT GTGTGCTC GGGCAGCCA AGTCTTTTCA CCGACCCCT 60  
CCGGCGCAG AAGACTGA GTTCATTTGA GAGCTGTAT TTGGGTTTA GCGGAGCTGC 120  
TGGGGCTTT GTCCCGGCC AGGACACAG TTAATTGCA CCGGGCGCG OCTGGCTTAT 180  
GATTTCTC AACCGAGG CGGCTCTGC CCTTACTCG TCCAGGCC ACTTGGCAGG 240  
CAGAGCCCT CCCAGCT TCGAGGCTGC TCGAGTGC CTGTTGGAAT GCACTAAGAG 300  
GACCTTTTG TGGACAGG TCTCTCAAC ACCTGCTGC TGGCTGCCAG GCAGGCCCTC 360  
TGGAGGCAA GGGCAGGAC TCACTAGAC CTCTCTGAG CTTCAGGCC AGGCAGTTGG 420  
CCGAGCCA AGCATTTGG TCTGTTGCC CGAAGGAGC AGAAGCTTC TTGGGCTCT 480  
TCCTTCTG GACAGGCC CCTGCTTGG CCTCACTAA ACTG 524

(2) INFORMATION FOR SEQ ID NO: 199:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 332 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 199:

45 GTGATACAG GAAGGTGAT CATCATCTGT CACATGCAA TTCTGTCCA CAGCTTTCT 60  
GTGATCCA CTCTGCTC TTGTGTATGT CCAATATCC CTAGCTTCT CCCCCTCTA 120  
GAAGGCTTC TTGATAGAT AGAATATAG AATGATGAC ATTCTCATG TGCATATAG 180  
AAGAGCCAC AAGCATGTC TTTAATATA AAGCAGATG TCAATCTTT TAGCTGCCA 240  
ATAGAGCTT GGTCTATCC TCTGAGCT AGGCTTAA AGGCTTCTG TTTTCTSAT 300  
TGTCTCAT GTTTGCCA GGTTTATTC G 332

## (2) INFORMATION FOR SEQ ID NO: 200:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 376 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 200:

5 CAGAGGAGCC CCGAGGCGTG TCGTGAATG ACATCAGTGC TTCCCTGAGC TGCCCTCCCC 60  
15 ACCCTCGGCG ATTATCCGAG GAACCTATG TTTTCTAGAA CCAAGAGGAG TGCTGGGACT 120  
CAGGAGCTGG TCGAGTAGG CTGAGTGSCA GCTCAGTCTCT AGAAGGTCTC TGAAGTCTG 180  
20 GACTGAGGAC CTGCTGACTC CCGAGGCGAG AGCCATCAG CAGAGGCTGC TGTAGGCGAC 240  
CTCCCTGTGG ATGCTGTGAC TCACACCAAG GCTGCGAGCC TCTGCGCTC ATTATAGGGA 300  
TTCTCATGAG CCGATGGGCG CTGAGGCGAG CCGATTAAAG CATCTGGCTC GTTTTITGGAA 360  
25 AAAAAAAAAA AAAAAA 376

## (2) INFORMATION FOR SEQ ID NO: 201:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1192 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 201:

40 CCGAGTAAAT TTCTATAGCA TTTATTTTAG TGAACCTATA ATGTTTCTTT GTATTAAAT 60  
ATTAGATTAT ATCTTTAGAT AATATGTTTA CTAAATTAAT AGGTAAATTA TATTTTATTC 120  
45 AAAAAATAAAT TGTGATCTTA ATGCTACCA ATTAATGTAC TTGTAGATGT ATCTTATCTT 180  
AATCTTGATC TTGTGTCGCC CTAAATCAGGT GTGAGGACT CTTCCTCCCT GGGGAGGTTT 240  
TTCTTTTTC GAGGAGGAGA GGGCTTTCCC AGTAAATGTG TCTAGATGTG TGGCGAGAAR 300  
50 AATCTGGGAC CACACGACAC CAGTTCTCTC CTTAATCCAC GTCAATTTGCC TTCTATCCCA 360  
GCTATGTTTC CAGTGTCTCT TGGGTGTTTC CAGAGCCAAC AAGAAAYGAA TAAATCTCTG 420  
55 KTGAGTTGTT TATTTGTTCT TCATCTGTT TTACTGTTA WTTCTGAGT TTATGGGTGT 480  
CTGTGAATTA AAAAAAATA GTGAATTA GTAAACTCA GTTGAGGA AATATACATA 540  
AATAGATATA AGCTGACCTG TAGATATAT CAGTTATTA TACCTTAGAG TTCTCTATGT 600  
60 TGTGTGCATA KTTTCTCTG ATCTTTCTGA TGGCGARCA AAAAAAGGAG TCATGTTTGT 660

WATGTGATTO GATGTGAGCC CGAGAGAGA GCATGCTGTG TTCTTGGGGA CAGGAAAGCT 720  
5 TGTGTGAGC AAGTCTTAAAC CACACCTTC ATGGAGATA GTTATATGTC TGGAGCATAT 780  
TTACAGCCGG CTTGCGAGTA AAGACTTATA GTTTGTGCA GTGGAAAGCG TCATCTTCCG 840  
CTAAGGAGCG GGTGTTGTG CAGCGGAAT GGTCACTGTC TGTAAAGCA CAGCTTCCAT 900  
10 CATTATGAT GCTGCTTACT CAAAGATGT GGTCCGAAAC AGCTTTTGG AGTGTCTCT 960  
TGATTCATG ATGAAAGCTG GAACATCTTG AGGACTGAT TAACTATAGG TCTTAAATA 1020  
15 ACTCTGACA CTTTTCCTT AGTTATCTC TACATGAGG GTTGTGAGA GCTGTGTAA 1080  
AGTCAATTT TCTGGGAAT ATTTCAGTG TTTATTTGCA CTTTACGCCA GTCTGTGTAG 1140  
20 CTTTATTTCT TCTAAACTCA CCAATATCT GATTAATAT CAATTTTAGG CG 1192

## (2) INFORMATION FOR SEQ ID NO: 202:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 589 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 202:

ATCTGGGCT ATCTTTGACA GCGGATCTTT CCAAGTTGAT GCTTCTTACA AGTGATATA 60  
35 GTCTGTCGCC AAGAGTGGAG AGCTTGAAT CTACAGAAAT TGAATGAAT GAGAGAAACA 120  
CACAGATTA GATTAACGAA CCAATGCTC CAGTTCTCA TTTCAAAATTT GTTCTGCTG 180  
40 AATATGATG TATCTGTATG AATCGAGAC AGATGTGTA AGTACAGCTG AGTCAGAAAT 240  
ATCAGAAAC AAGGAGAAAT GATACAGACA CCGGAGTGA CATTAGATTT TTAGCCACTG 300  
GTTTCGAGCG CAGAGAGAA AGCTTACAG AGATGTTTG TATTGATCTC ACTGTGAT 360  
45 CCGGGAATCA GCGATTTCCG TCACAGACTA CTGATCTGA GGCCTTTCT AGTGTGTTAG 420  
ATCAGAGGA AGCTATGGA ATTAAAGAAC ACGATCCAGA GAGGGGTCTC TCAGGTTCTG 480  
50 AGGTGAGAA AATCCCTGAG ACACCTTTG AAGATGAGG AGAGGAACTC AAGAGAGAAA 540  
AATAGAGAG TGTTCCTGTC CACCTTCTC TCACTGAAAC TCAGTCCCA 589

## (2) INFORMATION FOR SEQ ID NO: 203:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 847 base pairs  
(B) TYPE: nucleic acid



(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 203:

5 GCGACGAGCG CAGCTCTGCG GCGGCCGACG ACGCGTTGCG CAGGTTGCGG CTGAAAGACC 60  
GGAAGCTGCG GGAACAAAGTG AACTCTGATCG TGGACATCTC CAGATGACAC ATGATCTCTGT 120  
10 ATGACATGCA GCGAATCTCG AGCAGCTCAC ACCGCGGCGCT GAGGAAGACG ATGACAGACG 180  
TGGCGGGGAA GCTGATGACC CTGACTGAGC TGCTTNGAC TGCCCTGCGG CCGAGCAGCT 240  
15 TCCAGAACCC ACCGAGAGT CGAAGTAGCT GAGCCGACGA GAGGGAACCA GCTTACTTTTC 300  
CCGATTAAGT AUTGTGTGAC ATGCTCTCTG CCACTCTGGA CAGGCTTGAA CAAAGCAAGT 360  
CAGTGTGAG GACCAAGGCG GAGCTGCGTT GATGTGTTTG GTTCTGTGAT GCTCTCTGGA 420  
20 GGGAGGCTG GCTAAAGTGG GAGGCGCTTG GCGCAGCTGA GCGCCGAGGT GCGAAGATCG 480  
TCACCGCCAC TCTGATACC CTGATGAAA AGACTCTCAC TATGCTGCTA TGGAGAGCT 540  
CGAGCTCTCA GTTACAAAGTG CAGGCGACTG GAGGCGAGAC TCTTGGTCC CTGCGAAGA 600  
GCTTACTAGG GCGCCGAGTG CAGGATCTCG GAGGCTTCA GTTACCGCTG GCGCAAGTGA 660  
720 AGAAGCTGAT ATGAGGCTCG GCGCGGCGTG GAGGTGCGCG CCGCTGCTGG GACAACNAG 780  
30 AGGACAGAT TTATTCCGAG CTGCGAGAGG CAGCTGTGCG GAGGGAAGAA GTTATTACTCA 840  
CGAGCTCTG CTCTTATCTT TGTATTAAT GTTAAAGCCA GAAAAAANA AAAAAAANA 847  
35 AAAAAA

(2) INFORMATION FOR SEQ ID NO: 204:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 852 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 204:

40 ACAAATAC TGGCGAAG GATGTATG CTGCGCGCAG CATGAGCA ACNATGCGC 60  
50 GCGATACG GCTTCCGCA GTTGCGGCTG AAGACCGGA AGCTCCGGA ACAAGTGAAC 120  
TCCATGCTG ACATCTGCA GATGACATG ATCTGTATG ACTCTGACGA GATCTGAGC 180  
55 AGCTGACACC GCGCGCTGGA GAACAGATTT GACAGCTGCG CCGGGAAGCT GATGCGCTTG 240  
ACTGAGCTGC TTACACTGCG CCGCGGCGCG AGGAGCTTC CAGAACCGAG CAGAGCTCC 300  
60 AAGTAGCTGG ACCGACNAG GAGGAACAG GCTTACTTTC CAGTACTCA GGTGTGAGC 360

ATNCTCTCTT TCCACTCTCH TGNACCGAG CCTGAAAAA GCACTCTGAG TCCAGGACC 420  
AAGCGCGGCC CTGCTTTGGA GTGCGTTGCG TTGCTCATGCG CTGCTGAGCG GAGAGCTGCG 480  
5 TAAAGTGGK AGCGCTTGGC CAACTGAGCG CCGCAAGTGG GAAATGATTC ACCCGCACTC 540  
TCCATACCTT CATCAAAAAC ACTGTACTTA TGTGCTATG GAGAGCTCC AGCTCTCATG 600  
10 TACAAATGCA GCGGACTGGA GCGAGACTC CTGGTTCCTT GCGAAAGAGG GTACTAAGGG 660  
CCCGATCCA GGAATCTTGG AGGCTTCAAT TACCGCTGCG CAGCTGAG AGCTGGGTAT 720  
780 GAGGCTGGGG CCGGGCTGGA GGTGGCGCCC CCGTGTGGA CAAAGAGAG GACACATTT  
15 TCCAGAGCT GCAAGAGCA CCGTGTGGG AGGAGAGT GTTACTCAGC AGCTCTCTCT 840  
CTTACTTTTG TA 852

(2) INFORMATION FOR SEQ ID NO: 205:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1154 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 205:

30 GATTCGCA GAGGTTCTT GAGCAGAG AGTCTCTGCG CGGCTGCGC ACTGTTCTCT 60  
35 GAGGAGCTG GATTTGACT CATGTTGCTG CATGAGAGA TCTTGTGTT CCGGGGGTGC 120  
TGGCTGTCA ATGCGACTGG GCACTTTCA GACACACTTT GGTGATGCC CATCAGATTC 180  
CTGACATCG GCTATGCTGA CTTGTGCGG GGCACATAT GGGGCAAGAT CTTTTCCTG 240  
40 TCCACTGAG TCAATGCTGT CTGCTGACA GCGCTGCTCG TGGCTGTGT GCGCCGAGAG 300  
CTGAGTTTA ACAAGGGA GAAAGCTG CACACTTCA TGATGATAT CCAATATACC 360  
45 AAGAGATGA AGGATGCGC TCGCGGAGTG CTACAGAGG CCGGATGTT CTACAAAGAT 420  
ACTGAGGA AGGATGCTCA TCTTCCCGCG AGGATGAGC GCAATCTCT GCGCCGCAATC 480  
AAGGCTTTC GCGAGTGGG GCTGAAAC GGAAGCTCC GGAACAGAT GAACTGCTAG 540  
50 GTGAGATCT CCAAGATGA CATGATCTG TATGACTGCG AGCAATCTT GAGAGCTCA 600  
CAGCGCGGCC TGGGAACA GATTTGACG CTGCGGGGA AGCTGATGC CTTGACTGAG 660  
55 CTGCTTACA CTGCGCTGG GCGGAGGCG CTTCAGAAC CAGCCAGCA GTTCAGATAG 720  
CTGAGCCAC GAGGAGAAC CAGCTTACTT TCCCGATAT TCCCGATAT CTACAAAGAT 780  
CTGCACTCC TGATCCAG CCTGAAAAA GCACTCTGAG TCCAGGAGC AAGCGCGCC 840  
60 CTGCTTGA GGTGCTGCG TTGCTGATG CTGCTGAGG GAGAGCTGCG TAAAGTGGK 900

AGGCTTGGC CCACTGAGG CCGAGGTGG GAACTGGTC ACCCGACTC TGCATACCTT 960  
CATCAAAAC ACTCTACTA TGTCTGTATG GACGACTCC AGCTCTCACT TACAGTGGCA 1020  
GGGCACTGA GCGAGGACTC YTGCGTCCCT GGGAAAGAGG GTACTAGGGG CCGCGATCCA 1080  
GGATTCTGG AGGCTTCACT TACCGCTGGC CGAGCTGAAG ACTCGGGTAT GAGGCTGGGG 1140  
CGGGCTGGA GGTGGCGCCC CTGTGTGGGA CAAACAAGAG GACACATTT TTCCAGAGCT 1200  
CGAGAGACA CTTGTGGGG AGGAAGAGT GTACTCAAC AGCTCTGCT CTATCTTTTG 1260  
TATATATGT TAAAGCCAGA AAAAATATA AAAAAAATA AAAAAAATCG AGGGGGGCC 1320  
AGACCAATC TCCATATCT AAGKCCCN ANUN 1354

20

(2) INFORMATION FOR SEQ ID NO: 206:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1378 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 206:

TCCCGAGTG CAGACGAG GCGCTCTGT CTGAGGAGA ATTCAAGCT GGTGTGGAC 60  
TCAGCCCTTA GNCATTCOA AGCTTATG TTGTATCAT ATCTTACGTG TTGAGAGCT 120  
GACTGGAGA ACAATATGT CATTACAGY AATTTTATCT TAGAGATCTG TCGAGCTTAT 180  
TTCTGTGACA AAGTTATAT TGTCTANTAA GAGAGTCTTT AATGGCTCTCT GTGATATATG 240  
TAACTCACT TACAGGTGA CTTTATATG CATACATGA TTTGATGAAA GAGCTGAAA 300  
CAATCTGGG ATGTGTGGA AAGTTATCTT TCCGCTCTT TCGTGTGTC ATTGTCTTT 360  
GAGAAAGGA TGGCCCTGAT GCGACGAG GCGAGCTGT AATTAATAAA TAATTTCAC 420  
TATCAGACTA GCAAGCACT AGACTGGA AAGACGACAG AATACAGAGA ATCCAGCCCT 480  
TTCTATTTAC AGGTGACAA ACTGTATGA TGCATGTA TGTGTTTTT AAGCTGTAG 540  
CACTGTACA AATGTAAAT TTGCGTAT TACGAGATGC TGTGTGCTAT GAGAGACAC 600  
CGAGCAACT TGACTCCAG TGTGTGGCC TGTCTACACC AGACAACACA GAGCTGGGT 660  
CAATTTCCC TCAGTCTTT ACAAAGTTC CTGACAGGA AGTCTTTAC AAGGTGCTT 720  
TCTCGATAC TGAAGGTG AGTTTCTGA ACTGCACTGA TTTTATGCA GTTGAAAAA 780  
AAAAAGCT ATTCAAGA TTTCAGCTG TTCTAGACA TCTTCTATG GCTTTACTTC 840  
CTGAGGACA ATGTTTTTAC TTTATGATA ATTCAATGTT GCGAAGANT AAGGTGAGA 900

60

AACGCACT TTTATATAT AGTCTCTCT GAGAGAGAC TAAATTAGAA AGGAAAACT 960  
GTGCAATTT TCAATATCTC ATCTTAAAA AACCTATATC TTAATACAA AAGTCTCTTT 1020  
TGAATATAG TTACAGACA TGGCAGACG AGTTGTCTTT TAAATAGATA GTGCTATAC 1080  
TCATATATC GTTACTACAC TACTGCTTTT AAAAAAAA ACAGCTAT TTAATGAAA 1140  
CATGAGCAG GATTATAGT CTTTAAACA TATATTTTGT GACTTAAAA ATACATTTAA 1200  
ACTGCTCTT CTGCTCTAGT ACCTGCTTA GTCAAAATGA TTAATCTAT GTACACTGA 1260  
TCTTCTCTT TATTTTATA AATTATCAG AGTGAAGAA AAAAAAAA AAAAAAAA 1320  
AAAAAATA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA AAAAAAAA 1378

20

(2) INFORMATION FOR SEQ ID NO: 207:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1166 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

25

(xl) SEQUENCE DESCRIPTION: SEQ ID NO: 207:

AATCACTGC AATTAAACC CCGTCCCTC CAGAAAGTT CAGAACCGC CATGATGAC 60  
CCTCATTTA GATGGGCGC AATATTNAG ATGACTGAG GCGCCGABAG ACTGACCTT 120  
GAAGGGGGA CTCAGAGAA AGATCTTGA CATTGCCAA CATGTGGGC TTGTCCACA 180  
CATGTATGG GTCATGAG AATGGGCTT TCCAGAGCA ATACTTTAT GATAGGTGG 240  
ATGCTCTGA CCTGTGTGT CATGTCTTC GTGTGCACT ACTGACATG AGCCAGCCAC 300  
GCTGATGGC TGAACAGAT TCCAGAGCC TCGAAGTGT TGTGTGTGT AAGAGAGAG 360  
GGGCGCCAGA GCGCGCTTT TGAATCTTT GCTGTCTGA ACTGTGAGA CACTTGGAG 420  
TGATGTGCT CTATTTTCA ACTGCTCTG TTTTCTGTGA CATCTTGGAG GGGAGCTAG 480  
TCCGACACC ATCGCGGTG CTTAGGAAT GAAGAGTTC CCGGTCTGT CTCTCTCACT 540  
CTGCTCTCA ATGGGGAGG GAAGAAATG CTTTGTGGC TTTTGTACA CAGCTGATC 600  
GTGCGCTGG AAGGTGTCA CAGTGGCCC TGTGTGAGG ACTGTCCAG ACCTGTACA 660  
CCTTGTACC ATCAGGCTT TCTGGCTCT GATAGGTGT ACAAAGTGT GAAGAGGAG 720  
GAAGAGGCT TTTTGTACA GCAATTAAT TAAATAGAG GTCAATTCAC ATCTCTGTC 780  
TCTTGGCCAC CTTCCCTGT GCTTCACTGA CATATAGAT ACTGACTGC AATATCTTC 840  
ACCATTCCT GGAAGAGCT ACTTAGGGA ACAAATGT ATGCTATGT CCGATACAA 900  
GTAAAGATTT CCACTACA GCTGGGTGT TCTCTTTCT AAGTGAAGC CAGTGTATT 960

5 TCCGCCGAGT GTTCAGCTT GACCTCATCT ACTGATTTTT TCTAGTGTAT ATAGAGATGG 1020  
TTGGGCTTTT AAGGTCTACA GACTGTGGGC TTGGGCACCT GGGCCAGGG STTTTGTGGG 1080  
GGGCTTTGCC CTTAGGAA GTACCTTTTA GGGGCAAGA TTTGTGTGAT TTCCGCTTTA 1140  
CAGCTCTCAG CTGAGGGT TTAATA 1166

10

(2) INFORMATION FOR SEQ ID NO: 208:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 697 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 208:

20 TACTCTTAG ATTATAGGA ATTACATNTG AGATGACATT TCCATTTGAG AAGAAATA 60  
GTTCCTTTCA GTGCTTTTA TTGATTTCT GAGAGAGA GACTGGAGS ACATTCAC 120  
CCGAGGCTG ATATGAGT ATCTCAGA GGCAGAGCC AGCAGAAC AGCATCTA 180  
GAAAGTTACA ATTGGAAGT TTCTCCAG CTTCGGAT GACCTGCA AGCTGATCC 240  
AGAACTGCC AGGTAAATC TCTCATATC TGTCTATCC ACCACTTC AGCTGCCAA 300  
ATTACTAGT GCGTTGACT ATTCTCTTT ACCTTTATCA TTTAGGTGTA RCGATTCAC 360  
AATACTCTC GACTTTGCA TATAGAGCT GTGGTTCTGT GTGGTCCCTT GGTATAGAG 420  
CATCACCATT ATCTGGAAC ATGCAATATA TCGAGATTAT TCACTCTCTC CCGAGCTTC 480  
CTGATTTAGA AATTCACAG TTCTCAGGT GATCTATAC ATCTTAAAT TTGAGAGCA 540  
TTGATTAAG TTATGCAAT AAGAGAGT TAGTAGGGA TGGTGGCTA TCTTCTACA 600  
GTTCCTCTGT TAACAGAA GTGAGATC AGTTGATGAG ACATTAAT ATTATTTCT 660  
AATACTAATA AATATTAATA AATCTGAG GGGGGC 697

45

(2) INFORMATION FOR SEQ ID NO: 209:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

55

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 209:

60 CCGTAGTAC CTCCTATAG TGGGCTGGC CCGAGCCGGG GTGAGCTTGC CGAGGCTCC 60

5 GCTCCAGAA ACCATGTTCA AGGTAAATTA AAGTTCCTGT GGGCAGCCA GCTGAGCTT 120  
GCTCACTTC AAGTCTTGT CAGCACCAA AAGGAGTCA CTTCCGAAA ATTCCCTGA 180  
GGTTATGAG CTTTCACTCT ACTGATTTCC TGAAGGTCAA TCGAGTATG TGGAGAGCC 240  
AAGAGCCAG CTTGAGAAA GCATCTGCA GCTCCGACAT TATTCGAGC CATACACAC 300  
CTGGTGTGAG GAAGGTACT CCGAACTAA GCGCAGATG CAAGGTTTGG TTCAATGGG 360  
GTTAGAGAG TATGACTATC TCGAAATGC ACCTCTGGA TTTTTCGCA GACTTGTGT 420  
TATTTGTTT CTTGCTTTTA TTGACTCTT TTGGCTAGA GATTCAAAA TAAAGAGCT 480  
AGTATATCC CTTGTTTCA TGGGATGAG TGGTCTCTC TATATCCAC AACAGCCAT 540  
CGTATTCGC CAGGTCAATG GGGAGATTT ATATGACTGG GATTACAGG GATATATAGT 600  
CATAGAGAT TTCTGAGG AGACTTTCA AAGCCAGA ATGTGAGA ATTCACTGG 660  
AACTAGTAG AACTTTCAT GTCTGCCAT CTTATCAT TATGGTAAA CATTGGAAAC 720  
TCCATAGAT AATCACTAT TTCTACAGA AATGCGATA GAGTCACTA TTGATATAT 780  
TAAATGCTT TCTTTTCA GGAANAATA GACAGACT CTTTATCTT CTGTGAATC 840  
ATCTACAG CAACTAAC TGGATCTCT TCACTAGAG ATATGTACA AGCTTAGA 900  
CTCTCATTC TCAATTTCT ATTATGTAC CT 932

(2) INFORMATION FOR SEQ ID NO: 210:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(ix1) SEQUENCE DESCRIPTION: SEQ ID NO: 210:

45 GTCTATCTT AATTAAGGC TTCTCTTTT AAGCTTTTC AAGGAGCAG ACCACTTGA 60  
AGATTCGCC TGGGTTGAT ATGTGCTAA TTCAATTTAT AAAATTTAT CTGTCTTCA 120  
TTTTAGACT TTGCTATAT ATGAGAAAT GTCTTAATA ACAAATAT TTGTATTTAA 180  
TTTAGGAG ACTAAGGGA AGAANAACA AACTCACT TTTATGTAAG CTCAGAGAT 240  
ATTAAGCTT AAGGCTTTT TCTAGTTTA TGAAGATTG TACTACTAT TTTTATATAT 300  
TCTCTTTT GAGATGAACA GATCTCTGG GAATTTTGT AATTACATG GCATTTCACT 360  
GTGATCTCT TCAAGCTCAG ATGATTTCTA TAACCCATG ACAACTCTC TCTTTGTTT 420  
ACTCTCTGT GAATGTGAG CTCAGTTTC CCGAGATCG TGTGTTTATG ATGAGTCAGA 480

540 GTCCTTTTCC TCGCTGGAC AGTCTGGC CCTCTTAAT TTGGTATG TCTTCCAG  
600 TATCTAACC TCCAGTCTGA TCTGTATAG CTATCTAAC TGTAAATGT ATTATGAT  
660 5 ATGTGATTA TCTTCTTGA AGTTCTATC TTTTCATTT GATGAAATA AGTTTATT  
661 C

10 (2) INFORMATION FOR SEQ ID NO: 211:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 592 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

20 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 211:  
GAACTGCA TTGTAAACA CACTAACA GAGTACTTA CCTCTGAG ATTAAATA 60  
TAATGTTGA CATGATCAT GTACATCAT GGAATGACA GATCTTATG GTCTACATTT 120  
TCTTTATCC TGTATGATTT ACCTTCTTA ATCTTTGTT CTAAATATC TAAATCTTC 180  
TTCAATGTTT ATTCTTACT GACGAATGC TAACTTTCT TACACCTCG CAGAAAGGAG 240  
30 AGAAATGCT TTGGGGTGG GTAATTAAT TTGTGATGA AATATCAATA GATGAGATG 300  
GAAAGAGGA GACAAAGA GTTAAACA AAAACATG GTTTTTTAT GCTTTGACT 360  
35 GCTCTTTAA ATAGCTTACA AGCATATAC GTTAAATAC ACTTTTATG AATTAATAG 420  
TCCATACTA GTATGTCCTT CAAGAAGCA ATGTGCTTT AGTGCCCTTA GCTTAATTT 480  
TGTCTTTTG ACATGAGGA TTTTGTAGT ATTGCACTTA ATACCACTT ATTCTCAT 540  
40 ATGTATATT TTTTGCTAG CATTCTTT ACCACTAAC TTGTGATA GC 592

45 (2) INFORMATION FOR SEQ ID NO: 212:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 938 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 212:  
TGGAGTGGCT TTCCAGCTGA ATGAATCTTA TGTCTGGCT GAGAGTGGTT GTTTTCAT 60  
GTTCTTCTTA ATTTTCTCC TATGTGCTCT TGGAGTTTN CTTTGTATG TCTGTGTT 120  
GCCAGCTTT AATAAACA GGGCAACA AAAACATAG CATTCTGAA CAAATAGGGG 180

240 CCCACATGG ACCGATATG TCACTTTAAT GCACTTGAG AAAAATCTG AATCGAATA  
300 TCACTAGG AATGTATACT CCACTATTT TATCCATAT AATGTGTGT TTCTTTAAT  
360 5 TTGTCTTG TCGCAATG TGGCTTTCAA ATTAAATG CTTTCTCTC TTGAAACTT  
TTTCTTGA CTGTATAT TAAAGTTTG GAAGATTA TATTTTGA AGAGTTTC  
420 AACGAGA TACAAAGAG TCTCACTAT AATCTGTTT ATGTGCTTT ACAGCACT  
480 ACATTAGG AATATTAAT TACGAATTT AATATCTGT GTATGTTCT CTACTAATA  
540 AATATATC CTCACATA TCGGTCTG TCAATCTG CAATATCT CCAATCTCT  
600 TTATGAGG CACAGGAA CATTAGAT GTCTGCTTT CAGAAATCT CTCTATCT  
660 TCTGAGCC TTATTTTCT AATATCTGT TTCTTGTGA TTGTGTTAC CTCACAGC  
720 CATTGAGC ATGTGATCC CTCATTGCA TCAATCTAC CTGTCTTTA ATGTAAATA  
780 CATTTCAT GAAGATCTG ATGACTTCT AGCTTTTAT TTCTCTGTA AGCTCATGT 840  
900 GCTGAACA AACAGGCTT TTAATACCT GTTGAGGA AAACAAAA ATCTGTGT  
938 GGTGCTTT CCTGTAAA CTCATTAAA ATTCTTT

30 (2) INFORMATION FOR SEQ ID NO: 213:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1079 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 213:

40 AGCTTGGCG GAGATGATG GCATCTTAA GCTGTGCTGT GACTGTGTGT TGGGGAGGT 60  
GGAGCTGTT TTAACCTGTT GCGGCTCTC CTGTGCGGC GTGGCATCC CCGGGGAG 120  
TGAACCGCG GCGTCTCTC AGCTTCCAG TCAAGCAGC CTGGGGCGCG GCGCGCGCC 180  
45 CAGACACCC GAGAGTCTG TTCTTCCCTG GTTAGTGGA CTGTGAGCT GTCTCTGT 240  
GCTCAAGCC CTGTGGAGG TTGAAGCTA CTTGGGAGG TCGACACAG GCGGTGAGA 300  
50 GAGAGAGAA GGGCATGAG CCACTTCAG GATCTGTG TCCAACTCT ACACTAAG 360  
TTCCTCTCG CAACTCTGT GGCATTCGC TGGGGAGAT GTCTTTGAG TTCTACCA 420  
AGAAAGTCT TCGTGGCCA TTCTCAGAG AGTCATCCC ATGGAGATG TGGAGCTCA 480  
55 AGTGCATGT GTTACGCTG GCGAGGAGC AGAGCGCA CATCTGCGG GAGAGGTG 540  
GTGAGAACT CTGGAGAG ATCATCACTA TGTGGAGGT GATGAATCG CATGATGCT 600  
60 TCCCAAGAT GCGCAGAG TGGAGGTG ATAACTGTT TCAACAGC TTCGGAGG 660

TCAGCCGCTA CTTGATCAG AGTCTCTCC AGTCACTGTA TCCCTCTGGC ACTCTAGTGA 720  
CCACCACTAT GCGAGGCTC ATCAGAGACA CCGCTTGGCT CTGAGCGTGG CTGAGTCTCT 760  
GCGAGCTCTT TGAATGCTCC CTAAGCTTGG CTTTGGGAAA TTGCATCTTT GCGCTTTGG 840  
GCTCTGGAAC CTGCTCTGGG TCATTTGGTA GACTTGGGAG GCGAGCGCCC GCGTGGCTTC 900  
TTGCTTTTGT GGTTCGAGC CTGAGCTGAT CCTTTTATTC TTGCTCTAGG GTTCACTCTT 960  
GCGTCTACTG TCCTCTGATA GCGCTGGTGG GGTCCCGCTT CTTTCTCCAC TGTACAGAG 1020  
AGCCACCACT GCGATGGGGA ATAAATTTGA GAACHTGAGT TTGGCTGTGA AAAAAAAAA 1079

(1) INFORMATION FOR SEQ ID NO: 214:  
(A) LENGTH: 3791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 214:

TCAGAGGC GCTCTTGGCT GCGCGCGCCC CCGTGCATTC CTGTGAGGAA GCGCGCGCGC 60  
AGCCACCACT ATCTCTGGC ACTTACAGGA AGCGTTGGGC TCGCTGTGTA CCAACCGATT 120  
CGACCACTTA TTTACAGAG ATCTGAGCCC CTTGAGGCTG CTGAGGCGAG CAGAGAGAAA 180  
GAAAAAAAA GCGCGCGCGG GCGCGCTTGG GCGCGCTGGG GCGAGAGCG CHTCAGCGCC 240  
GCGCGCCAGA CCACTCTCAA CCGCGCGAGC AACAGCTGCG GCGAGGATTC CCGAGAGAG 300  
GCGAGAGCC CCGTGGCCCC CAGCTTGGC GTGCTTGTACA AGAAGAGGA GAGCGAGCGC 360  
CCGCTGGCGC TTTAGAGAG AGGATATAG AGGATTTGGA AGAGAGCTTG ATCAAGCACT 420  
TCAGCTGAAA GCGAAATATA TTGATAGAG ACCAGAGAG GAGACCACTC GTGAGCGAG 480  
ATTCGAGAG CCACTTGAG AAGAGGCTGA AGAGGCGAA TTTTCACTTG ATAGACCGAT 540  
TATTTAGCGA CCAATTTGAG GTCTGTGGTG TCTTGGAGA GGTGAGGCG GCGCTGAGCG 600  
TGGATTTGGC GAGGAGATG GATTTATTC TGTGGGAAA CTTGAAATTTG ATAGGCGATG 660  
TGGAGTGTAT AATCTCTCT TTTCACTATA CAGTGGCTTG AGCGAGAGG AGAAGCTGG 720  
AGGTAGCGGA TCTCAAGCT GCGAGCTGT CAAAGAGGA TTAAGTACT TCGATCATTC 780  
AATGTGACT GAGGAGAGC CTGAGTGA AGACATCAT CCACTGGCG AGACTGAAA 840  
TAGAGGAT GAGTTGAG AGTTAAGA GAGGCTTCCA AAGAGATGA CTTTGTGATGA 900  
GTGAGAGCT ATTCAATATA AGACCGGCG AAGATAGAA TTTATATTC GAAAGCAA 960

TCAGCTGCT GATGGGAT GGAAGAGGG ATTGTGTCTT CATTAATCAA AGAGTGAAGA 1020  
GGCTATGCT GAGATTTGG TATGAGCA TCATTTCCGG AGCCAGCAA ATGATATCAC 1080  
GTCTGAGCT GAGATCATTT TTGGAGACT TGGCGCGCCA GAGCTGGCG GAGGCGGAG 1140  
ACGAGTGA CAGTGGCGTG GTGGCGCCC AACCGTGGC AGCAGAGCG ACAGTCAAG 1200  
TGTCTGCT CTGATGTGG ATGACCGAGA GCGATTTCCA GCTCTGGCTT AACTGEMTC 1260  
CATTAAGCAA CCGTGGTTC TTGTGAGCC CTCTGTGTCA AACCTTTTGC ATCTTTAGG 1320  
ATTCCAGCG ACTAGAAAT TAAAAAATA AAGACTGTCA TTCTATCCAT TCACACCTAA 1380  
AGACTGANT TTATCTGTTT TAAAAATGA CTTCTCCCG TCACAGAG TAACAAATAT 1440  
GGATGCTGT TTGTATTTTA GAATGTATTT GATAGCGGG ATGTTTTCTA AATTTTTCAG 1500  
GATATGCT TCTTCAGAA TACTTTTCTA TTCTGCTTG CAATATGCA TTTTCCAACT 1560  
TCAATATAG GTGTGAGCAG TGTATACCG TTTAAGCTT TCACCTCAT TGTGTTTTTT 1620  
AATTAAGAT TTAGAGTTTC CCGCATTAC AACCTGTTTT TAATATTTGG ACATCTGCT 1680  
TTTAATACCT GCTTTGATA TTCAAGTG GTCACTGGG ACATGTTAAA CTTTCAATTTG 1740  
TCAATTTTTA TCTGTGTGG AATCTACT ATATGTATTT TACTTACTT TTATATTTTT 1800  
CATTTTTTGG GAAAAATCT TTTTCACTC TCATGAGGC TGTATATAT ATATCTTAAA 1860  
TCTTTATATA CAGAAATATC ACTACTGAA CAATTTCAAA GCGACTTGG TTTATTAACC 1920  
CTTCTCTCTT GATGCTGTA TTAGTTTCA ATTAATAGTG ATTATCATTT TCACTATAT 1980  
TTACTTTTTT AATCTTGAG TTTCCATTT TAAATCTTA ACTAGAGAT TTAATTTGCT 2040  
AAGTTCTTTT AACCTACTTA TTGTGTGAG GCGATCTGT TCAGTGTAG TAGTTTTATA 2100  
GZATGCGTT TTTTCTCCC CTTCAACAG GTGGGTGGA TAACTTGAAT TGGCGAATGT 2160  
GTATATTTTA ACTGTCTTG TAAATTAAT GTCTGGCGAT TTGTATGAT TCTCTGTGT 2220  
GAAGCTCCC AATTCAAA TGTACATCC ATATACAGCC ACCATTTAC CTTCTCTTGT 2280  
TCTAAGAGAA AACCGAAG GCGCTGTTG GTAGGTTGAG GTGGGGAGT ATTTTAATTT 2340  
TTGGAATTT GAGAGAGAG AGCTTTACTT TGTAGGTTG GAGCAGAGC ACTATACATG 2400  
AATTAATAC CAAGACCTT TACTGTCTT AATTTTCTTA GATGTATAT ATTGTGTGT 2460  
AAGTTGAGTA TTCCAGAGAA ACTGTATTT ATCTCTCTC TCTTCTCCA TTGAAATTT 2520  
AGGTAAATTA TGAATTTCTA TAATGGAGC ATCAAGCTT ATTAAGAGC ACATAGATG 2580  
ATCAATTA AAGTTTTTCT AGATGTCTT TTAATTTTGC CACATTTAT GATTAACGT 2640  
GAAGATTT TTAAGATTT TTAAAGAT GTTGTCTAG TCATTTTAT AATTTTCTA 2700  
CCTGTATAG GTATGTCCA GTTTAATA TATGGAGAT CTTCAATCTT AACATTTCT 2760

ATTAGCTGA TTGTTCTCA CATATCTTC TAAAGAAAC TTATTATGTTA TAAGGTATAC 2820  
TTTTTGATA AGATTATTA ATCTAGTTA CTTACTATTC TGCATTTTA GAGAGAGGT 2880  
5 AATGTTTTT AATGATGAT AACTGTGTC TGTGTTTGT GATCTATAGA TGTGAGCAT 2940  
GTTCGACT GTGTGATG TCTATATTA TTATTATTT ACACATATAC GTGTATCCA 3000  
GAGATTATTT TAGTCCATAT GATCTATGA CCAATGTTC ATTGAGACAG CACATATGC 3060  
10 ACTCTAAT CAGTGTGTT AGACTTTTCA AATATCTTAC TCAATTTCCA ACATATCCA 3120  
TGTATTATTA ACTCTTAT TTCCAGCAC ATACTATGA AAACACCTGC TACTCMAAC 3180  
15 ACACCTTC ATGTCTATCC ATTCTGTGC TGAAGACAA CATAGCAATA TCTGTATAT 3240  
TGCAGCTTT CAGATATCC TGAATTTAAA AAGTTGTGTC ATTATTTGTA TCTATGAT 3300  
ATAAATTTGC CTTCTAGTTC ACTTTGTGC AAGACTTAAA ACTGTGAAC TAACTTTTTC 3360  
20 TTATGTGTC GTATATCTG AAAATATAGA TTATTTTTCA TGTCTATTC TTAATATCTA 3420  
TAAATCAAT CAATATGAT CATCTTATTT GTCATGTGTT TCTGTGATTC TGACTGTGT 3480  
25 GCACACCTT GTGTGTTTAT AATTTTAAA TTGATTTTAA TATGGGTTT TTATTGCTA 3540  
AAACACGCG TTTTGATCA CATTTGAAA GGTATCTTAT CTTAATGACT AATGACTTAA 3600  
TTGGAAGT TGAATCTTG TAAATATCA AATCCAGGA CTTCTTGGG TTTATTTTAA 3660  
30 TTGTACTTC TTAGAGAGA TACATTTT TGTATATG AAGTTTAC ATACCAATG 3720  
CTACTTTTG TTGCAACG GGTCTATAG TCCGGGGGA AATCCCTTAC TGTATAGT 3780  
35 CCAATATG G 3791

40 (2) INFORMATION FOR SEQ ID NO: 215:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1334 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 215:

50 CAGTGTGCG TCTGTGTCG GCGCTGCGG CCGCGGCGT CCGCATGACC AGTATGCTG 60  
ACATCTTCT GTGGAGACGA CCGTATATGA CGAGAGCTG TATGCGCTCT GCTGTATGG 120  
TTACTGTGT ACCGACGCGG TGGCGCTGCG GGTGCGCTG GAAATCTGCG AGCAGACTG 180  
55 CCGCAGGCA CCGGTCTTAC AGAGAGAC CATGAGCAT TACGCACT TCCCATGCT 240  
CGAGGGCTG CTGATGCGC GCGCAGCT ACTGACACAG TCACTTCC AGATTCGCG 300  
60 CTTCCGCGAG GCATCTCTA TCGAGAGTA CTATGCTTT RATGAGGCT TTGTTGCGA 360

GGTGTGCGC AAGAGCTGT CMAAGCAC CAGAAAGAC CTGATATCA TCAGACGAA 420  
5 AACAGGATC ACCCTGAAA GTTGCAGAG ACAATTGAC AACTTTTAA CCGTCTTCA 480  
GGTGTGAGAG GAATGCGCG GTCCTCTGT GACATATTT CAGCAAGCT TCTCTCTTC 540  
TACCGGTTG GCGAGGAGT ATGAGCAT CTCTCTTTT CATTACACC GCTTTGAGAC 600  
10 AGGAGAGAA AACTGCAAT ATCTGAGCT CATTGACTTT GCTTCTGCG CTGAGCTAT 660  
GATCCMAAC TGAACCTTT GAGCGTTGTA CTCACAGAT GATGAGAT ACATGAGCTT 720  
15 AGCAGAGAA TTCTTCCAG ACTTGAAGGA GCTCAAGGTT GTATGCTG ACAGAGCTT 780  
TCTGAGCTG CAGAAAGCC TGTGTGAC TCTCTCTGCG GAAAGCTG GCGTCTTCTC 840  
TGAATGAAA GTCAGCTTCA AGAAGCTGTC CCGGGGCGTG GTGAAGTTG CCGCAAGCT 900  
20 GACCCCAAT AAGATATCA GAGAGCTT TGTGAGCTC GTGAGAGT TTGTGAAAC 960  
CTGCGCTCC GACACTGCG CACTGAAGGA GGTGCGGTT TCTCTGATC AGTATTCAG 1020  
25 GTCTGTGAC TCTCTGATG GCTTTGACA CAGGCGTCT GGAACGCTA CATGCGCAC 1080  
CTCGCGGCT GCTCTGCGG CTTGTATCAT GACTGAAGTT CTTCCMAAG CTGCGGCCAC 1140  
GCTGACATA AAGTTGCTCT GATTTGAGAG ACTGTGCTC GCTCCGCGGA GCAATGCGG 1200  
30 GCGTGAGA TGTGCTGTC TGTGTCTG AGCACTGCT GTCCCTGTAC AAGATGAT 1260  
GTATNCTG GCTCTTGGG AACTGAGCA TATCTGAGG AATGCTCT GTGCTGAGC 1320  
35 CATCCACAG AGA 1334

40 (2) INFORMATION FOR SEQ ID NO: 216:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1511 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 216:

50 GTGGGCGGA TCTGTGAGG GGTGCTGCG CCGCAGCGG GCGCGCTGC TACCTTCAG 60  
ACTGTGCGT ATGCTTCAA GCTGTTAC GCGCAGCTC GTGTATGCA CTTTCAGCG 120  
CAGAGCTGA TGGTGTGAC TGAATATAT CCGCCGAAAC CAGCATCCA CCGATCTGC 180  
55 CTGCAATTC CTGCAAGCC CCGCAGAGG GAGATAGGCT TATCTAGGCT TCTCCGCGG 240  
GAGATAGCA CAGTTTCCA GACACGGA ATGATAGCG TCTGCGAAA TGTGCTCTG 300  
60 ATGCGAGAG AAGGCTTCT TATGCGAAC CAGCTGCGA AACAGAT CTTATGAGG 360

5  
10  
15  
20  
25  
30  
35  
40

RTCTTCCCA ACCAGCTCTT GAGGCGCTTC CTGGAGGATT CCAGATACCA AATCTGCTG  
CCCCCTTTTG TGGGGCACA CATGCTGCTG GTACATGAG AGCCGAGGT CAGAGGATG  
GTACGAGTCT TAGAGTACTT GCGATCTCTG CCGCTCTGAG GTGGCTGCTT TGTATGACC  
ATCTCTGACA GGCAGGCGTT TATCTACTAC TCCAGCTCC CCAGCTGCTC CTGCTGGAG  
GGGAGCTTG TAGAGGCTT CAGCTGCTCT ACAGCCGAGA CCGACTGCTT GCTCAGCAC  
CAGCCCTTCC AGCTGACGAC CCTGTTGAC CAGTACATCA GAGAGCAGG CAGAGAGGAT  
TCTGTCTAT GTGCGCATG GAGGCCATG CTGTGACATG TTCCGAGCTC TTGCGCAGCC  
TGTTTAGCCA GCGCTGCGCA TAAATACAT CTGCGTATTT GCGTGTGCTC TCTCTAATG  
GACATGTGCA AGAATCTGGG GTGCGGAGT GTTGTGTGCA CTTCGTTTC ACTAGTATG  
AATATGTGAG GTATAGGCGC ACTTGAGAT GCHAGGATT CCAATTTGAG TGTGATGAC  
CGGCTGCTC CTATGTTTTC CCACTTGGG ACCTGATAG AGCAAGTCT CTCCATCTC  
CAGTCTGAG GCGAGATCC TGAAGATTA GCGCTATGTT CCGCTGCTC CTGCTGACT  
GCTGTGCTT GAGGCGCTC CTGAGCCAC CCGTTGCGGC AGAATCTGCC ACTGCCAG  
TAGCTGACC AGCAGTAT CTGTGAGAT GCHACTGCTG AGAGCTGCT GTGTGCGAG  
CTTGTGCTG AGTGTGTAC ATGTATAT TTCTTTACTG CTGAGCCAT TTGACCATT  
TGCAGAGAA GAGAGCAGA AATTAATG GTTGTGAG GTCATGAGT TAGTAAATG  
CAGACAGGG ACTTGAACA AGCGCTCTGC TGTGAGACC GCGTCTGCA TTCTTCACT  
AGAGCTTCT CATGAGTTA CCGAGATG GTTCCATCC ACATGCGG TGTGTTGCA  
TGTATTTCT CAGCTCTGA GTGTGACT GTGAAAGTT TGGGCACT GCTTATAT  
AATATGAAAT A

(2) INFORMATION FOR SEQ ID NO: 217:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 642 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 217:

55  
60

AGCGCTTACT TTCTCTGCA CAGAGATCT GAGGCGCGC TAGCTCTGAC TTGCGACTGT  
GACAAATGTC AGTATGAGG TGTAGGCAAA GACTGGGCAA TTGAGGAG GAGAGCGACC  
TGTGATCTG ACCTGAGGC GGGCGCTTTC ACCTTGGCTG GCTGTGCTT GCTCTTATG  
TTTGTGAGG TTGTGCTTGT TTGATCTCTT CAATGCTG ATAGGCTG GAGGCGGAT

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

ACCGGCTTG GAGTGTTC TTAACTCCA TCCATATAT AGGGCGCTG CATGTGTCTA  
GAGTAAAGC AGATGATG TGTTTTAGA CAGAGCTTG GAGCAGGCG TCTTACACT  
AATTTCTCT CTGTGTAGCT GACAAAGGT CTAAATTAGC TTACAAAGG AAGAGGCTGC  
CGTAGCCAG AGTCTGAGG GCAATGCTTT CAGTTTCCCT TGTGTAGAT TGTCTCTCAG  
TTCTATGAA AGCAGAGCC CTTAGGGGCG CTGGCAGAG AACACACCA TCTTAGGCT  
GAGCTGTGAA CAGCAGGGG TTGTGTGTCT GTTCTGTTC TGTCTTGGC GAACTTTCTC  
AATAAAGCT AATCTTAT TTATATTAC GTAGTGTCTG GG  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1241 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 218:  
GCTGCGCTG TTCCATTTTA TCTATATAGA TTTCAATCTA GGGCGCAGCC GTCTCTTAC  
TGATGTGTT CCGTTTACC CTGCGATGT ATATAGAT TTGTGTGAT GAAGAGCC  
AATTAAGCCA GATAGTCCC CAGGCGCTG ATATCATTA AGGCTTGGG AATGCAATAT  
GTAAATGCT TTAGTCTTTT TGTGTTTTA GAAAAAATA ACAGATGGG CTCAGATGCA  
TGGTACGTA AATATGCTC CTAGCTGTGT ACTCATTACT TTCTTTTGA TTGATATG  
AAGGAGCA GAGGAGAGG AATTAATG TCTCTTAGT ATCTCTGCA CTCAGTCTG  
ACTATGCGA TATTAAGCTA TATGAAATG CAGAGATTTG TATCTGAAC AAGGAGAG  
TTTGACAT TTATCTGC TTCAATTAC ATATTACTG AAGCAATTA ATAAACATAT  
GAATATCCA TTGCAGAG CAAAGCAC TAAAGCTTTT GTTCTTTTTT CTACATAGA  
GAATGATTT TTTTATTTT TTTTATGAG GAGCTATAT AATATGACC CAGTATGTC  
TTTTGTGAC TTAGCTTAT GATTCAGGT TACATGAG TTGATTTCTAG ATGTTACTA  
CCTGAAAG CATGTGCTG CTTATGTA CAGAGCGAG AGCTGCTG GAATAACAA  
AGCAATTA TCCAGACC AACTGTAGC TTATGTGCA GATGGAGT GTACAGAT  
CCCAATGTT GGGCTTTGG ATTTCCACAC CATGCCAGT GTGTGTGATC TTCTCTTTC  
ACATCTTGA TGAATTTG AATATGTA AATGAGCTT GAAATGCTT ATAGCATGAG  
CAGATCTTA TGACACATA ACAAATGTT CATATGTGA ATATTAGAA CTGTTACAC

CTCCAGTTAC CATATTTTC CATTTTGTG GATTCATAT TGAATAGCA GGGCTAGCA 1020  
ATTACTGCA AGTTTAGCC TGTGGTAAT ACCTTAGGT TATTAAATA TTGTAAATTT 1080  
TATTTAANTG TTCTGATG TTGGAAGCA ACAAATATAT CAGGATGCG TCTTTCCAT 1140  
GGGCTTATTT TTACCCCTT TTCTGTAAAG AAAAAGAAGC AATGCTTTAA TGTATTTTAA 1200  
AAGTTTTCG TATAGTTTCT AATTCATTT TTATTAAG T 1241

10

## (2) INFORMATION FOR SEQ ID NO: 219:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 219:

TGTTTATGTC ACTTAAGCA TACACACATG CACACACACA TACATATCA TTATTCAT 60  
CATTAAGTG GTGTTTCAG TGTCTGTG TCATGTTT TGCAGTTTTC ATTTCCAGT 120  
GAATTATGAG TCGAGGCCA CTTTCTAAC CAGATGTCT TTTCAGACA AGACCGGG 180  
RATTTAGGAA GAGTTGGA AGAGGGAGAG GCAAGGAAG AGGCTTTTAA ATTGAAGT 240  
TAATTTCTTA ACAGGAAGT GGGCTGATG ACTACATGT TATACCTCC ATCTTTGCA 300  
GGTGGGATG GAACATGCT TGTATCATC TGTGACGCT ATAAATCAT ATATCCACA 360  
AATACACAT CCATCCATCA ACATATCAT GTTTTGGAT GACACGCTA ATAGTTTGA 420  
GAGGAGTTT GTTCTTTT TTCTCTCAT ATACTCTTA ATGTTGTCA GTTATCAAC 480  
AAATATAGA CTAAATPAC ACATCTGCC AGTTTGTCT TACATGACA GTATGCTTAC 540  
CTGCCATTTA ATATTAGCT GTATTTTTC TCAGTATAT TTACCTGCA CTGTATTTG 600  
TTATTTAAC AGGAAAAA ACATCAAA AAGAAAAAT TACTGTAGC GCTTCATAT 660  
ACTATATAT TATATATAT ATTTGACAT TTGTGATAC TGTGAATTT TATCTCTGC 720  
ATATACTTTA TAGGAAGTA TTACGCTTA AAAATACGA ATTAATTTT ACAGGTTTC 780  
TGTTTGTCT GGAAGTAA TTGATGTGC TAGAATCAT GTTTTGTTT TTGGGTTTT 840  
TGTGTTTTTT TTTTAAATG TTACAGCAC TTTTTTTGT AATTTCACT TCCGAGTAT 900  
TGTACAGTT CACATGTT GTGATTTTG AATATGAGG AATATTTAA AAAAAAAA 960  
AATCCCGGG GGGGGCCCG TCCATGTG CCAAGGGGG GGTACGGCG 1020  
1080

60

## (2) INFORMATION FOR SEQ ID NO: 221:

- (1) SEQUENCE CHARACTERISTICS:

## (2) INFORMATION FOR SEQ ID NO: 220:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1258 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 220:

TGAATTGAG GCTTAAGAT AACATATGG GTTGGAGTT GTGTGTGAT AGGTTTTC 60  
TGCATTTTG ATTGAAGTT ATCCATATA ATTTTTCATA GTAAATTTT ATTAAGTAT 120  
AATATACATA TATTTCAAT GATTTTGTCT CTGAGGTTT TCAATGTGTC TGAATACAG 180  
ATAGTGGGC TTCAAGCTG GATTAATGCA GGGCAGGAGC GCTTACACTC TATGACAGA 240  
TTGTATATAC GGGATGCTC TGCCTGTGTT ATATGTGTT ACOTTTACCA TCCACTTACC 300  
TTGAGACCA GCGAGAGTG GAACAGGAC CTAGACAGCA AGCTCACACT ACCCAATGCA 360  
GAGCGGTGC CTTGCTGCT CTTGCCAC ACATGTGATC TGTCCCTTG GCGATGTGAC 420  
CGGAGACGA TTGACCGTT CAGTAAAGAG AACGTTTCA CAGGTTGGAC AGAAACATCA 480  
GTACAGACA ACAAAATAT TAATGAGCT ATGAGAGTCC TCAATTGAAA GATGATGGA 540  
AATTCAGAG AAGATATAT GTCTTTGTC ACCCAGGGG ACTACATCA TCTACAAAC 600  
AAGTCTCCA GCTGCTGCT CTGCTAGTAC TGTTTGTTT ATTTCCATC CAGTGTCTG 660  
GAGGCTTTT AAGTCTCTC GCTTTGGTT GCGACTGAC TAATTTTATA AGTACATTT 720  
AATGTCTCC TGAATCTGT CCAATAGCA GCGCATTTT CACTTAGAA AGACACTG 780  
AACCGAGTG CATTTCTCA TCTCTGAT TAGCCTTTTA CATTTGCTG TCTACATTA 840  
GTGCCATTA GTCCCTGCG TGTAGATCT TCTCATCAG CTTCAATTTG TGAATCGGA 900  
TTTGTGAGA AGCATGAGA ATCAGCACT GCGTTTAGA GATGTAAT CTCACTACT 960  
TCTGAGCTTA TTTTCCAT TGAATTTAT TGAATCATG ACTTCCATTT CAGAGGAAA 1020  
TGAGATCAA TGTATTTCC CAATTTCTT GTAGGCGTT GTTTCAGAT CTTCTGTCT 1080  
TGAATATTA ACATCTCAT CTGATGCA GAGGAGGGG TGTGGGCATC TGTGATTTT 1140  
TGCCTACTAG AAGTGTCCA GAATCACTG TATTTTGA AATTTCTAAC TCAATATTA 1200  
GTCTCTCTG TCTTGGCAT CAGATATAT TCCATTTT TGGGCCGGG CAGGCTGG 1258

55



- (A) LENGTH: 1693 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 221:

CACATATAT GAATATATAC CCTCTAAAA AGAGAAAAA AAATCAGCC GGTCAACTT 60  
AGAGCAACTT TGTCTATTA AGCAATGTT TATTCTACTA GAAAAATTT AATATCAAG 120  
ACTATATAT ACTATATAT TAGGAGTTC TTTTAAATAT GACTATTA ACAAATCTG 180  
AAACTTATAT CCACTATCA CCTCTTTAT TTTCTTTAA CACTCTGAA GCTTATGTT 240  
CTGCAATCA TGTGCAAT GTCATGCGCA GTAAATATAC AGAGAGATG TTTATATGCA 300  
ATTAAAGGCT GTTTGCACT TTAAAGACA CGTGGGCTGT AATGATTCCT GGGGGCAGAG 360  
TGGCATATG TTTTATCAA AATATGACAT ATGTCAATG TTTGCAATG TGTATGCTTG 420  
TTCAATTTT GAACGCGAG TTGACATAT ATAGAAAGTA TTACTTTCTT TCAATATGTT 480  
TTTGTCTAC TGGCTTACA GCTTTCTCAG AATATCTATG GGCACAGCAG CATACAGTT 540  
TCCATCTTAA TAGAATGAA ATTATTTTTC TATCTACTCA TACAGATATC TGGGTCAAT 600  
GAAAAAAT CATTTTATCC GTCTTTTAA TATATGTTTA AATATATAT TATATGCTCT 660  
GCAATATCA GAACAGCTCT GAGAGCAACA GTTTCCCAAT ACTCTTTCT GACCAATAT 720  
GCTGCAAGG TGGCTTCTC TTGCGAAGA GGAAGGGTG TGTGAATG GCTTACATATC 780  
TTCAATATAC CAATATGTA TACATTAAT TACATTAAT TTTATGCTCT CAGTATATTA 840  
TTATTTAAT TTTTATGTA TGGCTATCTC TTGGCTATAT AAGAAAGAA GCAATCAAT 900  
GAGATATCAG GATATTTTG TTTAATATCT TCCAGATATC ATGTTTTTAC AGTGGCTGC 960  
TATTTAGAAA AGTATTTCT CTATACACT TGTTTTACC TTTTGAACA TTGACAGAAA 1020  
TTATGCAATG GTTTGTTGAG ATAGGACTT GATGCTCTG TTTATCAAT TTTGCTTCAA 1080  
AGTGGCTTAC TCAAGAGCC CTAGACTGAG TAGAATTTA AAGATTTCA AAATCTTTCT 1140  
ATTCTCTTCT TAACTATCC ACAAATGAG GATTTGATA GCAATGATG GTATATGAA 1200  
GAATATTTCA CCAATTTCT TTTTATGAG TTTTATGCT TTTGATTTG AAATCATTTCT 1260  
TATTTCTTTT AAGATTTCT ATGTATGACT GTGAGATGC TACGCACT ATGCTGATAT 1320  
ATTCACTGTA AGTCTCTCTT CACCTGTATC AAGATTTGAG ATGCTCACT GATATATAT 1380  
ATTCTTTAG TAGAATGAG TTAATATATC AATGATCTT TAAAGAGATG ATGCAATCTT 1440  
GTATTTATG TGGCTATCT GGTCTATCT GAGGCAAT TAAAGATTT CATATGATAT 1500  
TTTCCATGAT TAAATATCA AACTGTACT TTAATATTT CTTTCAATC TGAATATCA 1560  
ATAGAGAGG CCAATATAT TGGCTCTTCA TCTTGAGAT TTCACTACT TTATTTTAA 1620

AGTTGTGAT AATTTTAAA ATCTGTGAA GAATAAAAAG TGAATTTAA TTAATAAAA 1680

5  
AAAAAAAAA AAA 1693

(2) INFORMATION FOR SEQ ID NO: 222:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 222:

AGCGTGGGT CGACCCAGC GTCCGCACT TGGCTGTGT GGAAGGGAG AAGATTTGT 60  
AAACCCCGA CGAGGTTCT GCTTACCGA GCGCCGCTCT GTCCGAGAC CCCCAGGTA 120  
AGCCACCTC ATCATGTCT ACAGAGGC AAACCTTCA ACTGAGACT TGGGGATTA 180  
GAAGAGGCT GAATATTA ACTCAAGT CATTTGACAG GATAGAGT AGATTCATT 240  
CAATGTAAA ATGACACAC ATCTGAAA ACTCAAGAA TCATACTCT AAAGACAGG 300  
TGTTCAGT AATCTACTA GATTCTCTT TGAAGGTGAG AATATCTCT ATATCATAC 360  
TCCAAAGAA CTGGAAATG AGAAGAGAA TGTGATGAA GTTTATCAG AACAAAGCG 420  
GGTCAATCA ACGTTTGA TATCTTTT ATTTTCTTC TTTTCTCTA ATCTTTTCT 480  
ATTTTAAA ATAGTTCTT TGTATGTG TTTTCAAAAC GGAATTTGAA ACTGCAACC 540  
CATCTCTTG AAACATCTG TAAATGAT TCTATGCTC ATATTTCAAT ATGTTTTCT 600  
TTCATGTGC TGAATTTTG TATCAAGCC TCAGTCCCT TCATATTACC CTCTCTTCT 660  
TAAATATAC GTTGCACAG AGAGGTACC TTTTCAAGA CATTCGATTT TCAGCTTGT 720  
GTCATTAAT AAGATGACC AATGAGTG TTCAATATGA CTTTCAATT GCGCTGATG 780  
TTCTAGCAT TGAATTTCT ACTCTGAC TGTGCTTTC AGTGGAGAT GGAATTTCT 840  
CAAGAGACT AACTGTGAA AATGACTT TCTTAACTT GAAGTACTT TTAATATTTG 900  
AGGCTCTGA CCAAGAGAG AGAATATCA GGTGAGTGC AAGATGAG ATAGGTGAG 960  
AGTATGACT AACTCAAG ATGCTTCA TCGAAGAA GCAATTTAG ATTTTAAA 1020  
AATCTGTCA GAAGATCCA GAAGTTCT AATTTCAAT AGCAATAT AAAGTATAC 1080  
ATGCAATAT GATACACA GAACTGCT CTTTGTGAT TATTTGAT TTTTGGCT 1140  
GGATATGG TTTTAAATG AATGTCTG TACCACTTC ATTAAATTA ACATA 1196

## (2) INFORMATION FOR SEQ ID NO: 223:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1791 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 223:

TCACGAGGCT GGCAGGAG GCTTCGAGCA GCTCCGAG TGACGAGG GCGCGCGCGC 60  
CGCGTTGCGC TGCAAGTCCA AGCTTCCAGG TAGCGCGCGG CAGAGGCTCA CCGAGGCTCT 120  
GGAGCTCTCG AGCGCAAGTC CCCCACTC AGTGCAGTCA TGAGTCCGGA AGTCAGAGTG 180  
ACAGGGGAGA ACCAGGAGCA ATTTCCTGCT CTAGCCAGT CCGCCAGGG GGCAGCGCTG 240  
GCCACACTCA TCCATCAGGT GCTCGAGGCG CTTGGTGTCT AGCTGTCTTG AGACTCTG 300  
GACATGCCCA ATTCTTAGAG GCTGCGTAG AGTGAATTTC CTTTACCTT CCGGCTGCTC 360  
ACAGTCTTTC CTATGCGGAC ATACGCTGAC TACTTAGCTG AAGCCCGGAA TCTTCTGCTCA 420  
CTACAGAGG CTCAGAGAA TAGGCTTCCA CACTCTTAG TTGTACCTCT GGTCTCTAAA 480  
GTAAAGCTTA TCCATATGCG AGTCTCTG GAGTCTCTGC CTTGCTTAT GTCCGCGAGC 540  
TGAGAGACT TGTATTTAG GCTCTGTATG CTGAGTGTCT TGTGCTCTC CTGACAGC 600  
GCAGCCAGG GCTCGAGGTT GACTTAGGCA TGGGCGGGA CATTCCAGGC CAGAGCTCA 660  
GTCCATTTGC CCGAACCTTG CAGGATGTGT GTTGTGGCTG TTAGGTCTTG CTGTAGGCA 720  
TTTAGGAGCA GTTAGGCGGT GCCAACCA ACAGAGGCA GAGCTGCGC CTGAGAGC 780  
AGATTGAGAG TAGGTGTGCC AACCTTAAAA AACCATTTAA AGTTACGAGC GCAGCAGGAG 840  
CGCGAGCCAC ATTTCAGGAC CTGTAGGCAC ACTGTACTGA GCTCAGGGAA CCACTCTG 900  
GCAGAACCA GCGCCAGCA GCAGAGAGC CTCAGAGGC AAGGGGCTTC GAGCGAGGCC 960  
CAGATTTTGG TCCAGTGA ATTGAAAGRA CTGTGTTTTC CTCCCTGCGG AGTGTGGGTC 1020  
CGAGTGGCTT GCGTGGCTCT TAGAGTCTCT CAGAGAGCT TCTGTGCGCC TGCGCAGCTG 1080  
ATTATCTTAG GTTCATGACC CTTCAGCTCC CTTAACCCA AACTATGATC AGAGCTTCTC 1140  
TAGGAGGAGG KAAATATGAG GTCATGTATTT TTCTGTATCT TTCTGTATTT TGTGACTTCA 1200  
TGTGTTGCAAT TGCTTCCGCG TGCCATGCTC TGTGCTGTGT TTCTCTTAGA GCTCAGCTTC 1260  
TGTCTCTGTT CATTAATGT CATTGATGAG GTGGGTAGCC CTGATGGGGG TCGCTCTCTC 1320  
TGTGAGCTTA CCGCAGGCG TTTTCTTCTG CAGCCATTC CTGCATGCTT GATGCCCAAT 1380  
TCTTATACCC TACCCCTGAC CTATTAGGCA GCTCTGAGG AGCCATAGGG CCCCACCTT 1440  
TACTCAGACC CTGAGATTC TGCGAGCGAG TCTGCGATGC CAGGATCTAC TGAGATGTTT 1500

5 CATCTAGAA TCTGTGACA CTACATCAT TTCTTTTCTT CTCTCTGCGC CTTCGCTCTCT 1560  
GGGATCTCTG CTGCTTCGAC CCGAGAGCTT AGGATGGCA GCGTTTCTTT AACATGTTGA 1620  
GAGATGATTC TTCTTGGCC CTGCGCATCT CCGAGGCTT GATGCGATC CTGCAAGGCT 1680  
TTATCTCTCT TTCTGAGGTT TGTTGGGGA GCGAGGGTTA TATAGATTTT ATTAAAAAA 1740  
10 AAAAGGTATA TATGATATA TCTATATATA ATATGAGCCA GAAATTAATC T 1791

15 (2) INFORMATION FOR SEQ ID NO: 224:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 224:

25 ACACTAGTGG ATCCAAAGAA TTGCGCAGAG CCGCAGCAGA TTGTTGAGCT TTCTCTGTG 60  
TGTGCGGCGC TCGAGCGAGC TGAGCTGTGTC CATCTCGAGG TAGCGAAGTG GTTCTTTGTA 120  
AAAAGAGCA TCCCATCACC GCTATAGAGA AATACACAAA CCGGTCTCTG GACCGGAGC 180  
CGATGAGTGG CAGCTTCCAC ATGATGCGGA ATGTTATCAC CTCAGAGCAG CCGATCTGC 240  
TCCGCTCTAG TGACAGCCCA TCATGAAAA AGAGAGGCA GCTGCTCTGC AGGTGTAAT 300  
35 CTGCTCTCTC CTCGACCCCG CTGCGCGAG TGAGCCCTGA CACATCTCTG AAGGACTCT 360  
TCAGTGTCTC AGGGGGCTCT TTGACCAAGC AGCCCAAGA TTTCAAAATC AAGCTTTTGA 420  
CAGGGGAGTT AGCGAGCCAG AATGTGGGCG AGAGAGGCTT GGTCTCTGTTT CCGCAGGCA 480  
AAGCTTATGA CCAATGGGCG ATCGGACTGG AGACCCCTGA TTGTGGGAG GATTGCGAG 540  
GATTAAGAGC TTCTCTACTG GATGGAGGCC GCGTTTCTCT GTTGTGTCTT GCGCTGTGCT 600  
45 CTCTCTCTCA GATTAACTTT TCTGTAGTA TTCTTCTTCA TCTCATGCGC AAGGTAGGCT 660  
TGTGTTTTHH AGTGTGTGCG TCCCGGAGCC TGAGCCGCAA GCTGATTTCTT TATCTGAAA 720  
TGTGACACTG AATCTCTCTG GTGGCTTTCT TTGCGCCCA TGCGATGCGC CTTGCGGCT 780  
GTCTGAGGGA CCGTCTTTTT TCGAGGGGCG GAGGGGCTGC CTTCTCTTGG TGTGTATTA 840  
GCTTTTCAA CATTGAGGG GATGAGAGC CTTGTGTGTC TGACGCGAGC CAGGTCTGCC 900  
55 TGAGAGCTGT GCGGCTCTCT TGTCTTTTCA GTGAGGTGTC CTGGGTGGGG AGGAGTCTC 960  
AGGCTCTTGG TCTCTCTCCC AGTGGTCCA GCGCTACTA GTGCGAGGG CAGGATGAG 1020  
CTGACCGCTT GCGAGAGTCT TACTTAGCTT CTGGCTTGG AGTCCGCTGT CGTCTCTGCC 1080

5 CAGAGGAAAGT TCTCCAGAGT TCACTCTTCC CTCTTCCCTG AGTTCTGCTG AATGCCCTAC  
CCCAGCTCTC TTTCCTCTCT GGTGTCTCTT GCTGGAGGG GCTGTGTCTG TGAAGCTCC  
CGGTCTCTAC CTGGCTGAG ACTTACAC ACCCTGGTTT TGTGTAGCG CAGCTCTCT  
TCTGTGTGG CCTTTGAAG GCTCAGCTC CCNTGTGCA GTCTTGGGT TTGAGCTTA  
10 TTGAATGA AGAGTCTAGT TTGTCTCG CTCTCCATTT CTGGCTCAG TTGTCTACAG  
GACGTGTCT AGGATCTCT GAGGCATAT ATCCAGCTGC CACGAGGGG CACTGTCTT  
TCCACTTAT GTAGTGAAC CCATCATCC ATGACCAAG GATATATTT CTGCTTGGC  
15 AGAGGAGAG GATCAGAGG AGCAGGGAG CTCTACCGAG CAGGTGTCTT CCCAGCATTA  
GGCCAGACA GTTGGAGCA AACTTCAGG CCAAGGCAGT CCTGATTA CAGGCCAGT  
GTGTCTACT AGTGTCTCC TGTGTGTGG GAGTGAAGG AATCCAGCT GGCAGGCTG  
20 GAGCCAGTT GGGAGCAGG TTCTGGAGC TCTCJAAAT CAGTAGCAG TGGTGGAAA  
GGCAGTCC GAGATACTC AGAGCTCC AGATTTCT TGAAGCTAG CCAGTGAATA  
25 AAACCAAGA CTGATTTTC CAGGGCTAG TCTCTAGGC AGGAGGACC CAGGATTTCA  
ACCAAGTTC AGTGTCCAG CTCTAGCT GCGCAGAGC GGAAGTCCA AGGAGGGGC  
CTGTGTGTC TCACTTCCA GTTCTTAAA GATGTCTCT TTTTATCTC CTAACTCTT  
30 CAGTGGGTG CAGCTCTC GTTAGCAGT GAGAGCATT CCTCCAGC TTTTCCCTC  
CTGGCCAGG AGAGATCCA GAAGCAGTA GGACTGTGTT TTTGAGGTAG TGGAGCGGG  
35 GGGCTACTG CTGGCACTGT GCTTAGGTA GGGATGTTA ATATCTCCG TGCATGCTT  
TATCTCTCT CTGATCCAA AGCAGTATC TTCTGTGTTT CAGCAGGTTT CATGAGTCC  
AGCTCAGCC AGTGGCCAT CTGAGCTGG TGTATAGGT GACCATCTGG TACNTTGAAG  
40 GGACTGTCT GCTCTCTCA CTCTATAGC ATCTATCTG GAGAGCGGG AGGAGAGGT  
GTTGGCTAG TCTGTGTCC TCTCCACTT CCTATGCTC TATGTACCC ATCTGTCTT  
45 CCTGTGAGA AGGAGAGAA GGGCATTA GAGATGAG GTGATATGT ATTACTTATC  
CATTTCTGA TAACTTTG TTATCTTAA AAAAAAAAAA AAAAAAAAAA AGGGGGG  
2517

(2) INFORMATION FOR SEQ ID NO: 225:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2424 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 225:

5 TTGTATCTAA TGAAGATTTG ATTCTATGGA CAGAGTCTTT CAGAGCTTTG CAGTGTATGT  
ATCAGAGAG TACAGCTTGC CATGTACTG GAGATTTAGT AGAAGTCTG TCAATATTTG  
10 TTTCGTCTTT GAGTCTTACA CCGCTTATC TTCAAGAAA AGATGTGAAA CAGCATTTAA  
TCCATGCGA GAGCGAATT GAAATTTGCC ATAACTGTT AACTCTCTTT AATCTCTATA  
15 GTCTCTCGA ACTTGAAGAT GCTGTATAG ATGTCTCAA GGAAGTGTGA CTTTTGAATC  
CCATGATTT TTTCATACT CTGZTCTCT TTCTACACA CAGCAATGT ACTTACCATC  
20 AGATTAATAT ACCATGTCT CTGGAAGTT ATTTCTCTG TCAAGAAAT ATCAGCTAA  
TAGAGGAAA AGCATATAT CCGCTCTGC GCTGTAACT CATATGTGC CTCTTGGCCA  
30 CAGTGTGGA ACCAGTTAG GCGAAGATG AGTTTATGA TGTATGTG CTAGACTACT  
TCTTTCTTA TCATAGTTC ATCCATCTAT TATGCGAGT TCCATCTAC TGTGAAGAT  
40 TACTGAAAC ATTAGTTAG CTGAGTCTC TACTGTGCTA TGAAGTTTG CCACTTCATC  
TTCACTGTT CCCAAACTT TGAAGTGGC TATGCGAGC TCAAGTCTGT ATCTCAGAAA  
25 ACTGATCAA GCTTTGTGT GAGTCTCTG TTTTGGAGA ATATATTTAA TGTATCTTAA  
TGAATGAAG AACTTTTTA AACACACA TTGTCTACAC GTTCTACGA CATTTCTCTC  
30 TAAAGTCCA AGTCTAAGT TTTTCTGAG CAACTGTGC CAATTTGATC AGCAGCTTAA  
TTACAACTT GATAAGCCG TATCAGACC TACAGTCTGA TTTCTCCAC CAGTGTGAAA  
35 TTTCAGAGC AGTGTCTCT TTAATGGGG ACCTGAGGGC ACTGCTTTTG CTCTGTGAG  
TACAGCTCC CAAACAGTTA AACCGCTC TAAITCCAC TCTGCAAGG CTTTTAAACA  
40 AATGAGAC TTGTCTCGAA CAGAGAACT CACTCAGCA GCAAGAGCC AAGAGAAAGA  
AAGCTAAGA TGAATGAGA GCACTCCA TTAAGGGGG GCTGTTTAGC AGTGAATGAG  
1200 AGCAGCTGT AGCAGCTGC ATCAGTGA CAAGAAAGA AACAGGGAG GTCTGTAGCC  
1260 CAGAGGAC TTCTGACAT GAGAGCAGG ACTGCTCAAT TATTCATCCA GGAAGTGAAC  
1320 AAGTCTTCC TTCCCTCGAA AATAGTCTG TTAAGGATA CCAATGAAA GTTCTATCTT  
1380 CATTTCAGA AGCATGTCA AATATGAGT CAGCAGTGC AGAGAGAGG TCCACATG  
1440 GTAGATAGA CAGTTTAAA GATTTTAAG ACCTCAGTG TTCCAGGAT TCTAGCTTAG  
1500 CCGAGAGA ATCTAGTTC CTTCTACTT CTATCTGTC AGTCTGTCT GACTTAGCTG  
1560 ACTTGAAG CTGTATGCG CAGCTTTGC CTTCCAGGA CCGTGAAGTT GCTTTATCTC  
1620 TCAATTTGG CCATTTCCGA GCACTCTTA GTCATATGCA GCAATGAGC ATTTTATATA  
1680 CCGTGTAG GACATGAA TCTACATCC ATGTCTCAC AAGGATATC TGGCAAGGA  
1740 AACAGGCTG CTCTGTGCA TTAGTGTAG CAGTCTACT TTTTAGTCCC TCAGGCCCA  
1800

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

CCCCCATCT GTTGTATATA GTTTGTGCTTA TTGTGTTTTG TGTCTCACTT TGTCCAGTCC 1860  
TCTCTCTCTG AATGCGAAGA TACATTTATA GCGTTAATTC TTGGTCAGGC AGACTCCAG 1920  
ATGAAAAA CTTGCATCTT CAGTACTT CTTAAGGCG AATCAGATA TGGATATGTT 1980  
TTATGTAAAT AAGAGTCAC TTATAGTGGT TTCAATTAAT ATGGCTGTCT GCGAGAGACA 2040  
GCGTTCCCTA GCGCTGTACA ATGTAAATTA AACTTACAGC ATTTTACTG TGTATGATAT 2100  
GGTGTCTCT GTGCCAGTT TGTACTTTAT AGAGCGAGAT TGCCTCCGAT CCGTGTGTT 2160  
CTTATATCA AATTATGTT TACTGTATTA CCGAGACACC ACGAGAAAT TGAATCTGTA 2220  
AAGATCTTC TTATGCTGTG CCGTGGCAT ATATTAATGG TCGTTTATTT AAGAGATAC 2280  
CTGTGGAGA AATTAAGCAC ACTGATGTA AAATTAATTT TTTTATTTTT ATGTACATGA 2340  
CTGATTAAT CTAATCTCT GCACTNAAT AAAGTATG TGTATGACTTA AAAAAAAA 2400  
AAAAAAAA AAAAAAAAA AAAA 2424

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1080 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 226:

ATATAGGAGC GATTAATCTGT TTACTATCTG TTCTTCTCGA TCCACTGACA AGCGGGTAC 60  
TAGGTGACAA GAAACAAAG ATCTTAATCA AAGAGGTCT TACAGCAACC CAAGTCTCA 120  
TCTTCCGATA GTTAAGATCA CCGCGCTTG AGTAAAGCTA CAGGCACAC CACTTCCGG 180  
TTTCTCTTGC GCGCTGTGCC AAGATGCGG ATGAGGCCAC GCGAGCTGTT GTGTCTGAA 240  
TCCCGGTGCT GAAAGCTAAC GCGCGACCCC GAGATCGTGA GTTGTGGGTG CAGGCACTGA 300  
AGAGGAATA TCAATGCTTT ATCCGTTATG TGGAGACAA CAGATGCTCT GAGAACGAT 360  
GATTTCGACT GGAATCCAAC AAGGAAGGAA CTCGGTGGTT TGGAAATGC TGGTATATCC 420  
ATGACTCTCT GAATATGAG TTGACATGCG AGTTTGACAT TCTATATACA TATCTTACTA 480  
CTCCGCCAGA AATTCAGTT CCGTAGCTGG ATGGAAAGAC AGCAAGATG TACGGGGGTG 540  
CGAAATATG CTTGAGGAT CATTTCAAC CTTTGTGGGC CAGGAATGTT CCGAATTTTG 600  
GACTAGCTCA TCTCATGCT CTGGGGCTGG GTCCATGGCT GGCATGTGAA ATCCCTGATC 660  
TGATTCAGAA GGGGTCTCT CAAACAAAG AGAATGCAA CGAATGAGA ATCAAGGCCAC 720

780  
840  
900  
960  
1020  
1080

TAGCGAGCG CAGAGGAGAC TTGTATAGCC TACATTAATA TTTTCTCTGG CATCACACTT  
AATCAATCA ACTGCTTCCC CCGACACGCT CACCTCTTAG TTATTACTAA GTATGTCAG 840  
5 TAGCAATTC TGGCGAAGAA ACAAACACAC ACCAAAGAT ACTCTACTT AGTTTCTAAG 900  
GCTGCACAG GAAGGGAAG ACTGGCTTT GGCACATCTA GAGTAAATTT ATATCCGCC 960  
CGAGTGGAG CACATCTCA TTCTGAGGCC AGCGGGTAA CTGAAAGTGA GTACATATAG 1020  
TCTTTCTGTT TTCTGAGAT AACCATCAA TAAAGCTGC TTCTCTCGG TAAAAAAG 1080

(2) INFORMATION FOR SEQ ID NO: 227:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1336 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 227:

TTGCATTCAC AATTAATCGG AGCGAGGCG AGGCAATTC ATCTGGGGG TGGCTGCATG 60  
GCTCCGASC TCTCTGGGTT TTGAGGATG GGTACASCT GCTTACGCT AGCAACATG 120  
30 TTATCTTGA TCTCTGGGTT TGAATCTGC AGCGGAGAC TGCCTATC AAAGATGCT 180  
GTGAATTCAC CAGAAACAT CTTGTAGATC ATCCAGGCA CTACTGAT GACTTGAAC 240  
35 AAGTGAAGTT TCTTTCCAA TGTTTTCCA TCTGGCAAT CTTCGCCAA GCATTAATAG 300  
ATCTCAAGG GTGGCTGCTT CTCTATCTGT CTTTCTGTT GGGCAATGAG ATCTTAAAG 360  
AATGTTTCCA GACAAATAG CTTGACCTTC TTATTCTCT CATGCAAT GCGAGCAACA 420  
40 AGTATGCGG CAGATGCGCC AGACCAATAC AGCTTGCAT GGCAGATCT GATGCAATTA 480  
ATGGATTAAC CCGTGACTC CAGATCAAT CTTCTGTCCA TGAATTCAG CAGCTTCTCA 540  
45 GTGAACAGCT TCTGCTCTC ATGTGATAA TGTCTAGGAC CTGGAAATTT GACTTCTGC 600  
AGTCTGAGG GACCAAGAG CTTCTCTTGG TCAAGCAATG GATCCAGGTC CCAATAGAG 660  
AGTGGCACG CTCGAGGGTT GCTCAGGCTC ATGTCTCTGC CCAATCTCT TCCGACGATA 720  
50 CTGAACCTTG ATGTCAAGT CAGTCAATGG GAGAGAGCT ATCCAGATTT CTGGAGAGCT 780  
ATAGAGGRC TGTATAGGTG CCGGGGAC TTCAATCTCC AGGGTTTCAG TTTTGGGCCA 840  
CACTGCTCC GGGCTGAGT TGGCAACT GCAATTTCCC ACACTGGCT GCGCCATGG 900  
55 AGAACATTT ATGTTACAGA AGGGAGGTT GTCTGTGATG GAACTATGTT GCTGCACTG 960  
ATCAGCTCA TCTTCTCAT CTTCTTATC CACATATTA TCTTCTCAT CCGAGGGAC 1020  
60 AGAGCCCTGT GATCTGGTT TATGATCCA SCCTTGGGC TGAAGGATGT CACACACTTG 1080

ATATATCTTC ACTGGTTCA TGGCACTC CTTGTGTGC ATCCATACAT CCAGTTTGA 1140  
TTCTCTGTC TTATGAGAG CAGAGCGATC CATTTAGCTG GGTCAAGGTC 1200  
ATCCAGCCTT TCTGTGACT TCTCTGTC TACAGCCGAG GCTTAAAA TGTATTTTC 1260  
CTCTTTCTGT TGAGGGGCTAT GCGGGGTGCG AGTTTCCAG GGAATCTGGA AGCATTAGA 1320  
GTCCCTGTGT AGCCAG 1336

15 (2) INFORMATION FOR SEQ ID NO: 228:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2043 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 228:

25 TCACTGATC CTTTCTTGT GTCTGAGGG ACCTGCTGCG GCGCTCTTCC TGGAGGCA 60  
GACCTCAGC CCGACGCGA CTCAGATCG AGACCTCTCC CTCCTCCCGG CAATCTCCT 120  
CCCTCTGCTT TCAAGCTGCG ACTTTGACA TGTCTACCCC CAGCAGATC CCACTGCGC 180  
CTCACTGCCC CTTCTCTGAG CTTCTTCCA AGACTCTCG GTCACTGCTT GCTGTCAAT 240  
CAGAGCGCCA GGTCTAGCA GCGCGGCGG AACGGTCTT GCTTSTTCT CCAATTAGCT 300  
CCAGTTGAGG TGTGAGACC GTGTGTGATA AGGTCTGAG CAGAGGCTT GCGCTCTGCC 360  
CAGGCTTGG TCTGTAGCAG CTGTTTTCG TCGAGGAGG TTGAGCGAG CAATGTCCTT 420  
CTCTGCGCTC AGGTCTAGCT GCGCAGCTG GCGCGATCC AGAGAGGCG GTGCGCTGTG 480  
GCTGTAGCTG TCGGAGCTG GCTTCTTAC CAGAAAGCC TCAAGCTTCC TCTGAGACA 540  
TCCCTCGCTC TGGCAGGG GGAAGGCTC CTTTAAAGGG TGTGCTTCC CAGTGGGAG 600  
CAGTCTGCGC CTGCCCCCTA CTAAAGCTC TGTCTGAG ACTTTTCCCC AAGTCTCTGT 660  
AACTGCTG AGGTGCGT CTGGCTGCGA GCGATCTCT GAGCAACTC TCTTGCCTCT 720  
TTTAAATTC ACTCATTTTG TAAAGCCA GCGAGCTGCT GTTTACTTNG CCGCTTNGCT 780  
TTTTTCATTT TTCTTTCCG TCTTTCTCT TGAATCTAGG GTTCAATAT GCTCTCTCC 840  
CTTGTGAGG GGAATGCTG CTTTCTTCC CCACTGCGG GCTGTTTCA GCAAGCCTCG 900  
NCGCAGCTG GGGGGCGGG ATGGGGGCTT CTCTCTCTCG GAGGGGTGCA GGTGCCCTC 960  
CAGGCTGCG AGGTTTCTTT CCGTAGCTCC CCAATCTGCC CCGCTGTGTA GAGTTGGGCT 1020  
TCTGTGCTTT GAACTCTCTT GCAATGCGA ACAGAGCTTT TCCAGCATTT GTTGTGTG 1080

TTTACTAC CTAACTTTA GAAATGAT GTTAGAGGT GCTTCCGAG GCGGACAGA 1140  
GTGTTTCTC GCGCTGAGA AGGCTCTGCT CAGGCTCTAG AGTCTCTTCC TCGCCACCG 1200  
ATACTGCGC TTAAAGAGG AAGCTGAGCG CACAGTCTCC AGAGCATTC GCGCCAGAA 1260  
GATGGGAGT TCTCTCTCC CTTTCTCTGT CTCTGTACC TCACCTAGCC TAGAGAGGAG 1320  
GTCAATTCAG GTTAGATTTG CTTCTCATTC AAGTCTCTCG GCTTTGGGT GGAATACAC 1380  
CAGCTTTGCG GCTGTGCGG AGACTCTCTC AGACAGGAA CCGCAGAGG AGACAGAGCC 1440  
TGCACATCC TCCACAGCCA GCGCTGCGC CAGGTGTATT GAACTGAGAA TTGTGCGACA 1500  
ACCAATTTCA TGTGCTGCG AACAGAGCC CAGAAAGCT GCGCTTCTCC CCAATGCGA 1560  
GCGCTGTCT CAGGCTCTTT GTCCCTTTGA CTTAGTGA TTCCACACAG GTGCCACAG 1620  
CTCTGAGCT TCAATTTCA TATATTTAGA GAGTTGAGA GTATATCAGA GATATTTTG 1680  
GAAGCAATT GGTCTATGCA ATGTCACTTT GGAATCTCT TGAAGTTTA ATGTTTTAT 1740  
TAGAGATTT AAGAAATTA AAGTCTTACA ATATCTTNG GTTTTTTTT TTCTCTTTT 1800  
ACCGACAA CTGACCAAT GCGATCTCTA TCAAGATGA GGTCTCTAAT GTTCTCTCT 1860  
GTCTTTAGGG AGGTATNAG CAGATGCGG AAGGGTGT TTTTTCTTTG ACTGCCCTCC 1920  
TTTCTACAG AATGTGCGA CCACTGCTG AGTGGCTGT GTTGTCTCT CTGTCCAGC 1980  
TTCTGTGTA GAAATATCA TTGTAGGGG AACTCAGGCT AGTGTACGG TCTGTGTTG 2040  
GGG 2043

## (2) INFORMATION FOR SEQ ID NO: 229:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 540 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 229:

TAAAGAG GCGGAGATC TGGGCTTCC TCTAGAGATC GATGGCTAG AGGAGAGCT 60  
GTCCAGTGT CCGAGAGCC TGGAGCGCTT GAACTCAGA CTCACAGCC GCGAGCTGAG 120  
CCAGAGGCC AGAGGTGCC TCGAGAGGA GAAAGACCG CTATGACA AAGCTCCAA 180  
CTAGAGAG GAACTGAGT TTCTTCCGA AGAGAGCCG AGAGCATTC TGTCTCTGT 240  
GCGCATTTT ATCTCTCTCA CCGTCTCTTA TGGCTACTCG ACCATGTAG CTTGCACTT 300  
CCCGAGACC AGCAGAGCT TCACTTGGC CCGTTGTCTA GATATAGCA GCGACTTCA 360  
GCTCAATAG GACCAAGGTG CTGGGTGTT CCGCTGCCA CTTAGTCTTC AAGCATGCT 420

TCCTGGGGC CGAGGCTTG CTTCTGCGG GTTCCGGTC TCAGAGGA 480  
CATGGCTG GTCCCTCCTT TACCCAGG GAGAGGCAAT AAGACACAA AGCTGGAA 540

5

## (2) INFORMATION FOR SEQ ID NO: 230:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 448 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 230:

AATTGTGAA TATTAGATA TTGTACTAT TTGACCCAC TCMAATCTC CATTGGAAA 60  
TACTGTGCA TACCACAGT ATTGTGAAA ATATCAGAT GCATATCAC AGCTGTGCA 120  
GACTGTAGT CCACTGGGC ATCAGGCA CAGCTAAAA TTGAAACAA AGATCTGAC 180  
AACAAGAG CAAAGGTGG GGTGAGAA GCTCTGAGT GTACTAGT CTGAATGCT 240  
ATGCACAGT GCCAGTGTG GTGTGATAT CCAGGAAAA CTGAGAGG CCCAGTCTT 300  
CACTCTGCT TGACATGAG CTCTGTGTA GCAGGAAGTG AAGCTTAGG CAGATTAG 360  
CTCTGAAGC ATTCCACAC ATACACAAA ATCTGCAAA GCATTAGCA ATCTGTGTA 420  
CTCTAAGTG TTCTGACCC AGGAACAA 448

35

## (2) INFORMATION FOR SEQ ID NO: 231:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 407 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

40

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 231:

GTATCTGCC CCAACCAAT ATGTGTGCT GCTTTNACC TGACTCTCC AATGTTGAC 60  
CCCAAGAGA GGCCTCTAGA CTAAGGAGG GCTGTGTGAC CAGGTGTGG TGGGGCTGCA 120  
TGAACTACC AGAGAGACAG AACTTCTGGA ACTGACCCCTG GGGATTCAG TGGATCTGCC 180  
TATGCTCTGG TCCACCCAG ACCTGTGAGA TTTTCTCAT GAGATGAC TTGTCTCTCT 240  
GCAGTATGTC CTCAGCTTC ATAGTACTC CCACAGCAC CAGCAATACA GTTACCTACC 300  
TGTGCTTGG GATCTGACC AGCATGCTG GAGAGGGG CACTGGGCA TTATCCCTTA 360  
ATCTGTATAC CAGGAGAGA CTCGAGAGT GAGACAGT AGGGACT 407

60

## (2) INFORMATION FOR SEQ ID NO: 232:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 830 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 232:

GTATTGATT TCAGGCTCT AATGGGCTC ATTAGCAAT CATCTCTCA TGTAGCAAT 60  
AAAAAAAA CTGANTACA TTCTTTCCG GATPACTAT AAGGAGACA TGTAGGCT 120  
ATTAATCAT CAGCACTCA GCACACGTT AGGTTGCCA GCGACAGCA TCACACATA 180  
GAGACAGT TATGTGCTC TCTCTCAG ACCGAGACA TGTATCAGT TTGTGTTTG 240  
GTTTATATT TTCTGTTAA AAAATGTGA AAGTTGTTT TTAGTAGAT GATATTTTA 300  
TAGCTCGAG TCTTTTGAA CTATAGAT GTCACTACT AAGACATA CCTTATTTT 360  
TTTTTTGTT TTTTTCAC TCAGTATAA TCAGAGAG TTAGGACC ATCTAGCTT 420  
TAGATCTCT TTTTATG TCTCTCAGG ATATGATGT TCCATACA GCACAAAA 480  
ACACAAAA ACATTTGTA ATATCACTT GATAGACTT AAGCACTGC TTACTTTGT 540  
GTCCAAAA TTATGTTGT ATATATAT ATATATAC ACGACACAC ATATATATC 600  
AACAAATA GAAATATA ACATCATT CACATTTGT CTTGCTGCT TACGATTTA 660  
ATACAGAAC TGTATGACA GTTAGGCTA TCTTAGATA TTTAAATTC AATTATAT 720  
AAACAGATT AACACACA ACAAAGCT CTATTTGAT GAGTATGTC TTCTATAT 780  
AATACTTGG CTTGGTTAT CCAATAGT CACATTTA CTGTTATCG 830

45

## (2) INFORMATION FOR SEQ ID NO: 233:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 932 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

50

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 233:

CGAGAGAAA GACCAATTA GATATGAAA CTCATATCAC TTCTATAT TCAGCTCTCT 60  
AGCAAAATG AACTTGGCC TAACTTAGG GATTAAGCAA TTTCTTTAT GTAGCAATG 120  
CTACGAAAC AAAAGAGTG AAGAGGCC TTTTATTATA CTTATGTAC ATATATGAC 180

180

5 TTTTGTGACA AGAATGCGAG AATATAGGCTT CATTTCTACG CTGCAAAATA ATCCAGATCT 240  
GCCTTCTAAA ATGATGTCAG TTTCATAGCT GAACATGCA AATATTATGC TGTGACTGAC 300  
TCTCGATTG GAGGAGGAG TACTTCTGTT TACAGAAAC YGATATTGTA TATATGTCAG 360  
GCTGTGATTT CTGACTATCA GCAATCTGCT GCAATGAC TTTTCTGCT CATGACTGT 420  
10 GGAATTTGA TACTTTTAA GCAATTTCTT CTATGAGCAC AGTGTCTCTT AGTGAATTT 480  
AATTTGACA AGGTGTGAT ATGCTTTCTT AAGCTGATTT GTATTAACT TCGAGAGCC 540  
TACATTTCT CATTAGGCTT RTGATGCTCA GTATCTTTCC AAGTGGCAGG CAGRGCTTNC 600  
15 CTTTCTGAT CAACATACG AATTTTTGA TTTCACACT ATGAGAGTC ACTTCTGAG 660  
TCCAAATTA AAATGCGA ACTGCTTTAT CCAAGATGC TGAATATAC TGTCTATCC 720  
20 AGTTTCTTA TACTTAAA GCAATTTTGT CTTTGTGTTG TTATGCTAG GCAATGGCGA 780  
GAAATGTGA TTAGCTGAG GCTTAAATC AGATGCAAT CTGTAACT GACCACTGTC 840  
TCAATACA GAGCTGCG AGCAATTTTC CAGAGCTGTG ATTGCCAGA ACACATATC 900  
25 CCAAGTTTC TAAATGAG CAATCTTAA AG 932

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2786 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 234:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 234:

40 TTAGCAGGCT GAGCTGTAAA AAGAGCAGAC ATCTGATC CCGTCTGCT TTATTCGCC 60  
CTGGTTTCA GAAAGGAGG ATATATGGG ACCACTGCC CTTCTTTGA TCCAGCATC 120  
45 TCAATGCC CCGAACCTT CCAATGCT CTCAATGCTG CTCACTGCT TCGAGCGAG 180  
CTCCAAATG GAGGGGCTT GCGCTTACA GTCTCTTTGA CTGTAGACA GCGCTCTGTA 240  
50 TCAATGAC GATGAGAAA GTCAGGCT AATGCGGAA ATTTGCACT TCAATGCTG 300  
TGTTTTGTG TTGTGAAAC TGAATGCTT TATTATTTAA CAGGAAGCTT GATTTTATT 360  
TTTGGAGTC TTGTGTTGA TTTTGTG GCGTGGGAGA GAGAGATTAG ATTTTTTGA 420  
55 CAGGAGTCC CTTCATAC AGTACTTGT AAGCGAGAC ATAGGTTGA AGAGGACAA 480  
CAGGCTGTC AATGATGTC TCTCAATGCT CTTTAAAGA AAGCTGTCC TCGAGCTTAA 540  
CAAACTACT ACAAATGCTT AATGCTTTT TCGAAGCTT TTAGGGAAG AATGTTAGT 600  
60

TCATGTTAC TAGATGCTC TTTCAGATTT TTACAGTGTT GAACTTAAAG AATTTTGACA 660  
GCGTGAAGAG GGTGTGTGAG AATCTAAATT ACAGATAGAT GATTTGTTCT TGTGAATTTG 720  
5 TTCTCTTTCC TTTTATTTTG TCGCTACCAT TTCTTACAT TTCCCTTGGG GCGCATCTCT 780  
GCGTCTTCC TTTTGTGTTT TGTCTTTGCT TTATGAGTTC ATTCAGCTC CCGTTHATG 840  
10 AAGGACACTG CTGTATGTA AGGAGAAAG TCTATGAGTC CTAAATTTT AAGTCMAAGA 900  
AAATGCTCT GTTTCCCTCT TATGTAACACT TCTGAAGAGG AAAAATCTCA ATAGCCAAAG 960  
TTAATATCC TATATATTA TTGCTTTGCG TTTCACCTAA ATTTCTGGG ATCACAATTT 1020  
15 CTTTGGGATA GAGGTGTGTT TGGGGAATAG ATTGCTTTAT CCGTCTCACT GGAAGAGAAA 1080  
GCTAGTGTTT TTGTACAGG TCACTAGGCC AGAAGGCCCA ATCTTATTT TCGCTCATCT 1140  
TCAAGTAAAG AGTAAATCTT ATCTGTGTG CCGTGAAGGC TAGAATGAA GCGTTACGCT 1200  
20 ATTCAATGTT TATGTGCGA ATGCAATGAT GCGCTGTGGA AGAATGAGG TTTTGTCTGA 1260  
AAAAAATA AGACAGATTT GTGTGTGCA AACTGCGCTT ATCAATTTT TCAAGGAATTT 1320  
25 CTTTCTTCCC AAAAGAGGA GTACAAATTT GTCAATTTCTG AAGAGGCTT ACTTTATACC 1380  
AATGATGTC AGCAATTTGG ATGCCAGGGA ACGAGAGTG AGACACTTAC ATCAACCAT 1440  
CTCAATGCG CAAATGTTTC TTGTACAGAT GTTGCAGATTT TCGCATTTCT CCATATATAG 1500  
30 GGAATGAAA ATGAAATAAA GATGAGAGG ATGTAGATA TCGTTTCTG CCAAGATGCT 1560  
TTGAGTGA CTTTGTGATA TTGACTAGAT TTGAAATATC AAGATGATTT AGATGAATCT 1620  
35 ACAAATAGT TGTCTGTC TCAAGTCCCT TTTCACATTT TTGACTACTT AGCATCTATA 1680  
TTCCACACTT AGCTTTTGTG TCACTATAT CCGTTGTCTC CATTAAATTC ATTTGCAAGT 1740  
40 GTTATGCTC AATATTTTAA GCGACTTACA CAAGAGGAAA CTGCATTTT AAAAATCTTT 1800  
CTGAGATGG AGAATATGTA TTCTCTTTTC CTATAGCGCT CTCCAAACA AAAAACAAT 1860  
AGTAAATCTT ACTATTTAGA AACTGCTGCT ACTTTTCTTT TTCTTTTAGG GGTCAAGGAC 1920  
45 CCGCTTTTAA GCTACATTT GCTTACATA AATATTTGA GCAATTTGCA ATACTAAAT 1980  
ATTTTATTA GACTTTTAT TTTCCTTTT CATTAAGGGA TCGTGTATAG TAGAGTTGCT 2040  
50 GTATTAATAC TATCTAGCC GTTTCCCTGC TTTCCTTTCT CCGTCAATAG CCGTATGTC 2100  
CTTTCAGGA GCTCTTTTAA TCTTAAGATT CTACATTTCA TCGTCTTAT CAAATCTCTT 2160  
TACCTTTTAA ATACTGTC CCACTGATA TTTCATCTTT CAATGCTG TCTTAATATC 2220  
55 TGAATCTGTA GTTGAATAC AGCTATTTAA TATTTCTGGG AGATGTGAT CCGCTCTCTT 2280  
KGTGCTGCC CAAGTTGTT TTGCTTACT GAGACTCTTT GATATGCTTC AGAGATTTTA 2340  
GCGAATCACT GCGCATGCC GTGGAGTAC TGGAGTAAA ATTAATATAT CAGGTATAG 2400  
60

ACTAGATCC ACATGAGCA CTTCAGCTC CTCTTACTT GTTTGGGAA AAGATATAT 2460  
GAGTGTACTA CCATCTAAC TAGATTAAT ATAGTCTGAT TGTTCGAAT ACATTTT 2520  
5 TCTGCTTGG TGTTTTCCC ACTTCCAT GTACTGAGA AATTTGACA ATGTATATG 2580  
ATCAATTAA AATATTTAT TTCTTAAG CCCTTTTCC CTGTGTAT GTCCAGGACC 2640  
CTTCTCTTT CATGGAGAG ACAGTATTT ACTGATAT AGTTTGAAA GTTATGTAA 2700  
10 AAGAAATTA TATTAAGG GATCTTTC TTTCATAT TTGTATTT CTATTATG 2760  
GTAGGCATA TTAATAATA ATATG 2786

15

(2) INFORMATION FOR SEQ ID NO: 235:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 458 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 235:

GGGTGCAGG ATTCGCAG AGAATGTT TGAATTTCT TCTATTTTA AGATCTTCT 60  
30 CTCTCTTGA TGTTGAAAC TTACTTACT GAGATGCT TTCAACATC TGTGTCTTT 120  
TACTCGATA ATCAAGCTA TGCATCTAT CAAGTGTAT ATCTGTGGA TAGTTTAT 180  
35 GAGTCACA ACNAACACT AGATTAAG AGATGATCA AACCTGCAC CCTGCAGAA 240  
GAATTTCTT TAGTACTGT TCMAAGAA GGAAGGAACT TTTTATACA AAGAGAGAA 300  
TTCTTTTAA ATATGAACC TGAATTCAG CCTTCAGTA CAGGTACTT GCGCTCTTC 360  
40 TTTCATGCC ATCTCTTTC CACTCTCAG TGGAAATAT TTTAAGTGT TTATATCAT 420  
AAGTCTTGT GAACCTAAC AAGATATCC CTTCCTAA 458

## (2) INFORMATION FOR SEQ ID NO: 236:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 591 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 236:

AGAGTAAGA GGAATATAT TCTTGATTC CTTCAGGA AATCTCTTC TATACTTTA 60  
AAGCTCTGT TCTTTCTAG GATCCAAAT TGTGATTC TCTTAACAT CAGGTACAA 120

TTAGGTGCT AGATTTGTA TGAAGGTTA ACTGAGTCA AATTTACTT GATCTCTCTG 180  
AATATATCT GCAGCTGCA ATGAGAGAG AACAGAAA TGTATGTGA TGTCTCTGCC 240  
5 CAATCTATC ATGGGTTTT GATTTGTTT GATPATTTT TCTTTTTC TTATCCCTCC 300  
TTATAGCC TTGTGACAT TGGATATCC CAGGCACTC TCCACATCA ATGTAACTCC 360  
ATGCAATG CTGCTCTCG TGTGTATC TAAATTTT GATAGGAAA CAATATCTTT 420  
10 TGAATAAAA TAATATACA AACATAAA GTTATATG CCAATTTA GCTTGGAAAG 480  
TTTTTTCAT ATGCGGAG AGATTAATCT TTTTAAAG TACATATAT GAACTATAT 540  
15 GTAACTGTA ATATTTGTA AAGTCTGAT TTCCAACT CTTTGGAT T 591

## (2) INFORMATION FOR SEQ ID NO: 237:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1286 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 237:

TCCTTTTAA GTACAGAG GAGAACTCG AACTCAGAG AAGAACTG CCTTCCATC 60  
TACAAAGCT GAGTTACTT CTCTCTTC TTGTCTAG ACTGGCTTC CAGCAGCAG 120  
30 GAGTACTT GGGCAATG ATGTATATG TCAGATATA ACTATCAGC GAGTAAAGG 180  
CAGGCACTG GCATATGTA AGACAACT ACAGTCTTT TGTAAAGAT CTCTACTTA 240  
AATAGAAC AATTTTACA AACCACTCC GTTTTCTCT CAGGAACTC CTCCACTCA 300  
40 CCTTCACT CTTCATTT TTCCACTCC TCGACTCTC AGACTCTCT CACTCTGAT 360  
TCCAGAGG GGTTTTCTC CTCACAGG CAGTCCACT CATCTCTTA TACCAAAAT 420  
45 AGAATGTA CATCTCTG GTTATATAG TGTCTCTCA CTGCTTTC CATATGCCA 480  
TCCATATAG AACATATAG ATACAGGAA TATCCAGAA GAGTTATCA CAGTCAAGA 540  
GCAATCAG AATAGAGA AGGATATGA GAAGAGAC AAGGAGAAA AGGAAACCC 600  
50 AGGATATCT CTCTCAGG TAAATATGA CTTGCGATG AAGTGTAGA AGGATATAT 660  
CAGGAGAC ACATACAGA AATATCTAA AGAGCAAG AGGAAAGA ACCGGGAGT 720  
780 GAGCTTCCC CTGAACAGA GAGCAGGAA CACTACACT CAGATATGC ATGTTTCTG 780  
840 CTTTCTCT ATATATATC CAGATATGA TACTATAT CTGTATTT TTCTGATTA 840  
900 TTTTAAAGA ATTTACTTA ATCTCTTC TGTGTGTAG TATGAAGAT TAACTTTTT 900  
60 TCCAAATTA AAGATGAT TTTTCACT TTTTGTCT GTACTATTC 960



AAAAAAAAA AGAGAGCAT GACTTATAT CCAGAGAGG AATCGAATG AGTCACTTA 1020  
5 CAGGGAATCT AAGAGACTT GTTACTCTG TACTATACA GATTATCTGA GAAAGATCA 1080  
AGGTTTCAC TTGGCCACA GTTTTATCT TAATCAACA CCACTCTCTT AAGAGCTTC 1140  
ATCAGAAAG GCACCAAGG GGGCTCTCTT AGGCTTTGA GATTAAAC TACTCTTAT 1200  
CCATTACTG TGTGAACTT CTGGCTTAG TATTTTAGG GGGATCTT ACTTTTWT 1260  
GGTTTTCAC ACTTTTTCG TTGGC 1286

## (2) INFORMATION FOR SEQ ID NO: 238:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 734 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 238:

ATGCGACGC AGAGAGCA GCGAAGAT GCGAGGCGG AAGGCTAG CGGACGAC 60  
CTCTGCGCA AGCTATTC CTCCGTGCA GCGGCGAGT GCTTGAGAG GCGCGCGCG 120  
ACCATCGGC CCGAGACAC CTTCGTGAC CAGCAGGCT TCTACGCGC CGCAGACCTG 180  
GGAGACTGT CGACGCGCT GTTAGCGAC GTGAGTACT ACGAGGAA CTATCTTTC 240  
GTGTCTCTG GCGTCATCT GTACTGTGT GTACGTGCC CTATGTGCT GTTGCTCTG 300  
GCTGTCTTT TCGGCGCTG TTACATCTT CTATCTGCG ACTTGAGAT CGAAGCTTGT 360  
CCTCTTTGC CGAAGATGA GCGACGCA TCATATGCT TGGCTGAGG CATCTGCTTC 420  
CCCTTCTCT GCTTGGCTG TCGGCGCTG GCGCTCTCT GGTGTCTGG AGCGACCTTG 480  
GTGTCATCG GCTCCACGC TCGCTTCAC CAGATTGAG CTGTGAGCG GAGGAGCTG 540  
45 CAGATGAGC CCGTGTGAG TGTCTCTCG GACTGCGCG CTTCCGCGCG CAGCTGCGCC 600  
ACCTCTGCC ATGCTGTTC TCGAGGCTT GCTCTCGCG CCGACAGCG CTTCCATCA 660  
50 CAGCGCGCG GAGGATGCC GCGTTGAA ATTAAGCTT TATGCTGTG ATTCAAAAA 720  
AAAAAAAAA AAA 734

## (2) INFORMATION FOR SEQ ID NO: 239:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 809 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 239:

5 CGGGCTTTC AGGATACCG GCTGTTTACA CGAGCTCTAC CCTCAGAC GCAATCATGG 60  
CAGCGGAAA GAGACAGAG AAGATTCGG AGCGGAGAG GCTGAGCGG AGGACCTTC 120  
10 TCGGAAAGT GATTCCCTCC GGTGAGGCC GCGATGTGCT GAGCGCGCG CCGCGACCA 180  
TTGCGCGCT GAGCAGCTTC GTGACAGAG AGGCTTCTC ACGGCCCGC AACCTGGAG 240  
15 AGCTGTCCA GCGGCTCTTA CGCAGCTGG AGTACTTACA GAGCACTAT GTTTCCTGT 300  
TTCTGGGCT CATCTCTTAC TGTGTGTGA GTTCCCTTAT GTTCTGCTG GCTCTGCTG 360  
TTTCTTTCG GCGCTTTTAC ATTCTTATC TCGCAGCTT GAGTCTCAG CTGTGCTCT 420  
20 TTGCGGAGA GTTGAAGCA GCGATCATG ATGCTCTGCG TCGAGGCATC TCTTTCCTCT 480  
TCTCTGCTG GCTGTGTGG GGTGCGCGG TCTTCTGCT GTTGGAGCC ACCCTGTGG 540  
TCACTGCTC CGACGCTGCC TTTCACAGA TTGAGGCTGT GAGCGGGAG GAGCTGACA 600  
25 TCGAACCGT GTGAGGTATC TTCTGGAGC TCGCGGCTC CCGGCGAGC TCGCGAGCC 660  
CTGCGATGC CTGTCTCTGA GCGCTCTGT GCTTGGGCGC ACAGGCGCT CCTATCAAA 720  
30 GCGCGGAGG GATTCGCGC TTTCAGATA AGCTGTATT GGTGTCTAT CAGGAAAAA 780  
AAAAAAAAA AAAAAAAAAA AAAAAA 809

## (2) INFORMATION FOR SEQ ID NO: 240:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2201 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 240:

TCGACGAGG GGTCCCGCAA CATGGCGGT GCGTGTGTC AGCGCGCGG CTGACGACA 60  
GCAATGAGG CGGCTCTCTA CCGCGGTGA GGGTGGCTT CCGCTTGGG TGTGCGCTC 120  
10 TTGAGCGCG TCTGTCTCCC GACATCACT GTATTCGCA CTGCGCTCT GCGCTGTGT 180  
TTTACTAGA CCGGAGAGG TCGAGGCGC GCGTTCCTT CTCAGTGTG CTCTGTGCTT 240  
55 CAGGCGAGC TCCCGCTCTC CGGCGCATCT TCTTTCCTCT GTTTCGCTC CATCTCTCT 300  
TCTCAGCTT CTTCCCTCT CAGCGGAGT AKCCGAGCA CCGGCTAGT CTTGGCAGCC 360  
CGGCTTTCG AGGTCTCAG GCGACCGCA AGTTCGAGA GCGCGAGTC CCGAACGAG 420

480  
540  
600  
660  
720  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440  
1500  
1560  
1620  
1680  
1740  
1800  
1860  
1920  
1980  
2040  
2100  
2160  
2201

ARGGGCTGCA GCGATGATAC GGGCTGTATC CCGGCGGTTC GGGCTCTGATA GTGGGGGTTC  
TGTGCTGATC GATCTCTGCG AACTATATCG ACCGGTTCAC GGTGGCTGGC GTCTTTCCCG  
5 ACATCGAGCA GTTCTTCAC ATCGGGGACA GTAGCTCTCG GCTCATTCGAC ACCGTGTCA  
TCTCCAGTTA CATTGTTGCG GATCTGTCTT TGTGCTACT GGTGTGACGCG TACATCGGA  
AGTATCTCAT GTGGGGGGC ATTGGCTTCT GGTCCCTTGT GACACTGCGG TCAATCTTCA  
10 TCCCGGAGCA GATTTCTCG CTGTCTCTCC TGACCCGCGG CTTGGTGGCG GTGGGGGAGC  
CGAGTTATTC CACATCGCG CCGACTCTCA TTGGCGAGCT CTATTGTGGC GACACGGGGA  
CCGATATCTC AGCATCTTCT ACTTTGGCAT TCCGGTGGCG AGTGTCTCG GTACATTTTC  
15 AGGCTCCAAA GTGAGGATTA TGGCTGGAGA CTGGCACTCG GCTCTGAGCG TGACACCGCG  
TCTAGGAGTG GTGGCGGTTC TGTGCTGTCT TGTGCTGTCT TCTCTGTGACA GTCTCATCTT  
TGGAGTATTC ACTGCTGGA CCGGAGTCTT GGTGTGGGCG CTGGTGTGCG AGATCAAGCG  
30 CCGGCTGGCG CACTCGAAC CCGGGGTGTA TCCCTGTGTC TGTGCGACTG GCTCTCTGGG  
CTCTGACACC TTCTCTTTCC TGTGCTTTTC CTGGGCGGCT GGTAGCATCG TGGCCACTTA  
35 TATTTTTATC TTGATTTGAG AGAGCTCTCT GTCCATGAC TGGGCGATCG TGGCGGAT  
TCTGCTGTAC GTGGGTATCC TTAACCGAGC CTCCACCGCC GAGGCGTTTC AGATGTGCT  
GTCCGACTCG CTGGGTGATG CTGGGAGGCC CTACCTCAT TGGCTGTATCT CTGACCGGCT  
CGCGCGGAC TGGCGCGGCT CTCTTTTTC CGAGTTCCCG GCTCTGCTAT TCTGCTGTAT  
GCTCTGCGCG TTGTTTGGCG CACTGGGCGG CCGACTTTTC TGGGCAAGCG CATCTCTAT  
45 GAGGCGGACC GCGCGCGGCG ACAGCTGAC GTGCAAGGCG TGTGTGACGA AGCAGGGTTC  
ACAGAGACC GATTTGTGCT GCGCGAGCG GCGCGCTGCA CCGCGGTGCC CTTGGCTAT  
GTGCTATCT GAGAGCTGCG GCTCTACCTA CTTGCAATC CTGCAAGCT GCGCGTGGCG  
50 CCAACCGAGC GAGGCTCG GCTTACCGCC TTGGGCTGGC CCACTTTCCA GAGGAGCGCT  
GCGCGCTGCG CGAGCTGCA GACACTAT GGTAGCTCA GGGGAGAGCG TGCGGGTCCA  
GAGGCGGAT CTTCTTCAC AGGGGAGGCG CCGAGGCTC GGTGTATTT GTACGAGAT  
55 AAAAAATTTA GCGAGAGGCC AGGTGCTGCG TCTGCTTTT CTCTGGGTG CTTCTGTATCT  
TGCACCGGCT GTTACAGCCA GGGGCTCTGA AGACTGTGG T

## (2) INFORMATION FOR SEQ ID NO: 241:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1661 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 241:

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

GTCTCTCGCG ACATCGAGCA GTTCTTCAC ATCGGGGACA GTAGCTCTCG GCTCTATCAG  
ACCGTTTCA TCTCAGTTA CATTGTTGCG GATCTGTCTT TGTGCTACT GGTGTGACGCG  
TACATCGGA GATTTGTGCT GTGGGGGGC ATTGGCTTCT GGTCCCTTGT GACACTGCGG  
20 TCAATCTTCA TCCCGGAGCA GATTTCTCG CTGTCTCTCC TGACCCGCGG CTTGGTGGCG  
GTGGGGGAGC CGAGTTATTC CACATCGCG CCGACTCTCA TTGGCGAGCT CTATTGTGGC  
GACACGGGGA GCGAGGCTC AGCATCTTCT ACTTTGGCAT TCCGGTGGCG AGTGTCTCG  
25 GGTATTTTC AGGCTCCAAA GTGAGGATTA TGGCTGGAGA CTGGCACTCG GCTCTGAGCG  
TGACACCGCG TCTAGGAGTG GTGGCGGTTC TGTGCTGTCT TGTGCTGTCT CTGATTTATG CCGGAGCGCG  
30 GAGGGGAGC CTTGGAGGCG CACTCAGATT TGCACAGGCT GAGGCCGACC TGTGTGTGGG  
CAGATTTGAG GCTCTGCGCA AGAATTTCTA GTTGTGCTCT GTCTTTCCCG GCTTTCTATG  
35 CTGTGGGCTT TGTGAGGGC TCCCTGTGTC TGTGGGCTCC GGTATTTCTG CTGGTTTCCC  
GGTGTGTCT TGGGAGACC CGACCTGCC TTCCCGGAGA CTCTGTCTCT TCTCTGTGACA  
GTCTCATCTT TGAATCATC ACTTGGCTGA CCGGAGTCTT GGTGTGGGCG CTTGGTGTGCG  
780  
840  
900  
960  
1020  
1080  
1140  
1200  
1260  
1320  
1380  
1440

GTCTCATCTT TGAATCATC ACTTGGCTGA CCGGAGTCTT GGTGTGGGCG CTTGGTGTGCG  
AGATGAGCGG CCGGTTCCCG CACTCGAAC CCGGGGTGTA TCCCTGTGTC TGTGCGACTG  
GCTCTGCGCG TGTGTGGCG TTGTTTGGCG CACTGGGCGG CCGACTTTTC TGGGCAAGCG  
TGGCGGATTA TATTTTTATC TTGATTTGAG AGAGCTCTCT GTCCATGAC TGGGCGATCG  
TGGCGGAT TCTGCTGTAC GTGGGTATCC TTAACCGAGC CTCCACCGCC GAGGCGTTTC  
AGATGAGCT GTCCGACTCG CTGGGTGATG CTGGGAGGCC CTACCTCAT TGGCTGTATCT  
AGATGAGCT GTCCGACTCG CTGGGTGATG CTGGGAGGCC CTACCTCAT TGGCTGTATCT  
50 CTGACCGGCT GCGCGCGGAC TGGCGCGGCT CTCTTTTTC CGAGTTCCCG GCTCTGCTAT  
CTGATCTCAT GCTCTGCGCG TTGTTTGGCG CACTGGGCGG CCGACTTTTC TGGGCAAGCG  
TGTGTGACGA AGCAGGGTTC  
55 CATCTTAT TGAAGCGGACC GCGCGCGGCG ACAGCTGAC GTGCAAGGCG TGTGTGACGA  
AGCAGGGTTC ACAGAGGACC GATTTGTGCT GCGCGAGCG GCGCGCTGCA CCGCGGTGCC  
CCTTGGCTAT GAGAGCTGCG GCTCTACCTA CTTGCAATC CTGCAAGCT GCGCGTGGCG  
60 KCGCTTGGCG CCGGCCGAGC AGGGGAGGCG GGTAGCTCG TGGGCTGGCG CCGACTTCCA

CAATTG

1146

5

GAGGACCT GGGCGTGTG CAGCTGCCA GACACTACT GGTAGCTCA GGGAGGAGG 1500  
TGGGGTCCA GAGGGGGAT CCGCTTCAC AGGGGACCC CGAAGGGTC GGTCTATTT 1560  
GTACGGAT AATTTTGTG GCGAGCCC AGTGGCTGC TCTCTCTTT CTCTGGGTG 1620  
CCTCTGACT TCCAGCCCT CTTCACCA GGGCTCTCA A 1661

10

## (2) INFORMATION FOR SEQ ID NO: 242:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1146 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 242:

15

60

AATCCAGGC TCTGTGCGA GGGGTGAG GCGAGGGGC GCGAGGGCG CAGTTGCGA

120

CATGCTCAG AGCGAGAGC GCGAAGACC GTTGGCCAG CCCAGCGAC TTGACAAAC

180

CTTTCAACA CCACAGGCT ATGAGGCTCC AGCGCTGCC CATTGGCTC CAGCCTCAG

240

TCTCTCTTG CAGCCTCTGA GAAGGCTCAG CCCACAGAA CTTAGAACT ATGCTCTATA

300

CAGCACTCAG GCTTCAGCTG CAGCAGCCAC AGCTAGCTG CTGAGAGAC AGGAGAGCT

360

CAACCGGAG GAGAGGAGT TCGACCTAAG GAGAGGAG CTGACGATG CTGCTCTGG

420

RGGACAGCT ACTCAACAG AATATTGGC CCGTCTACT TCTTTTGTG CAGTTGAGC

480

CTGCTTTTC CAGGAGCTT CAGTGAAT CCCACAGAA TTTCAGAGA CTGTATCCAC

540

CATGTACTC CTCTGAGT GACAGAGCT GGTCTCTTC CTGAACTTC TCGCTGCTT

600

GGCAGCTTC TGTGTGAAA CGAACATGG CGCAGGCTT GGGCTTTCTA TCTCTGAGT

660

CTCTCTTTC ACTGCTGCT CTTTGTCTG CTGTACCGC CCAATGTATA AGGCTTTGCG

720

GAGTGAAGT TCTTCTAAT TCTTCTTTC CTTCTGAT TCTTTGCTCC AGGATGCTT

780

CTTTTCTTC CAGGCGATG GTATCCGAG TTGGGATTC AGTGGCTGA TCTCTGCTT

840

GTTGTGCGG AGGCGACAC AGCATATTC GTCTCATTC TCTGTGCTCC CCGTCTTTC

900

ACTGGCATG CTGTGCTAGG AATGTCTAT GTGAAAGGA TCCACTCTT ATACGCGCC

960

ACAGTGGCA GCTTTGAGA GCGCCAGCA GATTTGCTG CTGTGTCTT CTCACAGCT

1020

GGGTGGCAA CCGCACTTG CCAATGAGC CGCTGGGCT GCTGAAGATG CTTTCCGGC

1080

CCCTGAGCC CTGACTGGA TGGCTGAGC CTGCTACTG AGGAGCTGA CTTAGCTCC

1140

GTCTCTAGG TCTCTGGAC TTGAGAGAC ATCACTAAT GATGCTCTT CCGTATGCT

1200

CCCAATCTA TGGCATGAC TGTGAGCTT GACAGGCTG TGGGAGTTC ACTGTGAGT

1260

AGTCCGCCA TCAAGCCACA CTCTGACAG CTCTGACAG CCCACAGCA CTTTCTCTT

1320

GCTGTGAGC GGTCTGCTT TGGTTATTT AATTAAGAG AATGTGAGC TGGAAAAA

1380

AAAAAAAAA AAAAAAAG GGGGTCGC

5

## (2) INFORMATION FOR SEQ ID NO: 243:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1350 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 243:

15

60

AATCCAGGC TCTGTGCGA GGGGTGAG GCGAGGGGC GCGAGGGCG CAGTTGCGA

120

CATGCTCAG AGCGAGAGC GCGAAGACC GTTGGCCAG CCCAGCGAC TTGACAAAC

180

CTTTCAACA CCACAGGCT ATGAGGCTCC AGCGCTGCC CATTGGCTC CAGCCTCAG

240

TCTCTCTTG CAGCCTCTGA GAAGGCTCAG CCCACAGAA CTTAGAACT ATGCTCTATA

300

CAGCACTCAG GCTTCAGCTG CAGCAGCCAC AGCTAGCTG CTGAGAGAC AGGAGAGCT

360

CAACCGGAG GAGAGGAGT TCGACCTAAG GAGAGGAG CTGACGATG CTGCTCTGG

420

RGGACAGCT ACTCAACAG AATATTGGC CCGTCTACT TCTTTTGTG CAGTTGAGC

480

CTGCTTTTC CAGGAGCTT CAGTGAAT CCCACAGAA TTTCAGAGA CTGTATCCAC

540

CATGTACTC CTCTGAGT GACAGAGCT GGTCTCTTC CTGAACTTC TCGCTGCTT

600

GGCAGCTTC TGTGTGAAA CGAACATGG CGCAGGCTT GGGCTTTCTA TCTCTGAGT

660

CTCTCTTTC ACTGCTGCT CTTTGTCTG CTGTACCGC CCAATGTATA AGGCTTTGCG

720

GAGTGAAGT TCTTCTAAT TCTTCTTTC CTTCTGAT TCTTTGCTCC AGGATGCTT

780

CTTTTCTTC CAGGCGATG GTATCCGAG TTGGGATTC AGTGGCTGA TCTCTGCTT

840

GTTGTGCGG AGGCGACAC AGCATATTC GTCTCATTC TCTGTGCTCC CCGTCTTTC

900

ACTGGCATG CTGTGCTAGG AATGTCTAT GTGAAAGGA TCCACTCTT ATACGCGCC

960

ACAGTGGCA GCTTTGAGA GCGCCAGCA GATTTGCTG CTGTGTCTT CTCACAGCT

1020

GGGTGGCAA CCGCACTTG CCAATGAGC CGCTGGGCT GCTGAAGATG CTTTCCGGC

1080

CCCTGAGCC CTGACTGGA TGGCTGAGC CTGCTACTG AGGAGCTGA CTTAGCTCC

1140

GTCTCTAGG TCTCTGGAC TTGAGAGAC ATCACTAAT GATGCTCTT CCGTATGCT

1200

CCCAATCTA TGGCATGAC TGTGAGCTT GACAGGCTG TGGGAGTTC ACTGTGAGT

1260

AGTCCGCCA TCAAGCCACA CTCTGACAG CTCTGACAG CCCACAGCA CTTTCTCTT

1320

GCTGTGAGC GGTCTGCTT TGGTTATTT AATTAAGAG AATGTGAGC TGGAAAAA

1380

AAAAAAAAA AAAAAAAG GGGGTCGC

5 (2) INFORMATION FOR SEQ ID NO: 244:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1529 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(2) INFORMATION FOR SEQ ID NO: 245:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1537 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

15

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 244:

15 TCCGAGAGCC CGGGGGTTC CAGCTCTGCC TGTAGCAGAG CCGTATGAGG GAGGAGAGG 60  
AGGATGTGCT GAAATAGCTC CGGAGATCT TTTTCAGCTA GGGCATMAAC TGTGCACTGA 120  
ACTGTCTGCC GAGAGCAGCT GAGGAGAGC TGAAGTTCCA CTGGTGTCTC TGGGGCCGCC 180  
GCCCTGTGGGA ATGGGGCTCT CTGTGCTCTT ACCTTTGTGC CTTCCTTGGC CTGGCAGATT 240  
CAGCTCAGCC CAGAGGCCCC TGGAGCTCC GGGCCCTTGG GTGCCCTTCT GAGTGTGGGG 300  
AAGGCGAGAC TCAGATGAG ATCCCATTTG ACTCCCTCTG TATGTACTGT GCCCTCTGCT 360  
GGCTCTTGAG GCTCTGGAGT CCGAATTGTC TGTGTTAGTC AGTGACAGG TTTCCAGGGA 420  
ATATATGTCA TGTGTTGGTC CAGCTACTG GAACCAAGA GACAGTACTT TGCAGAGAA 480  
AGATCACTG CAGAGCTGAC TGGATGCT AGATTTAGT CCGCATGAT TCTTCTGAG 540  
GAGGAAGCT GTTTTCAAAA TAGTTTCAT CATGAGTCTA TCATGAGCT CCCAGCTCTC 600  
CAGCCAGCT AGAAGGCAA CAGAGTGGC AGATTCTCT GCCCTGTGTC GAGGTTGAG 660  
GCCAGATGT ATTAGCTGCT AAGCCAGCA GGCCTCTCTC GCGAGCTGC TACCTTGTCT 720  
TCACTGTATC CTTTGCTCCC GGGGCTGAG CTATTAAGCA AAGGGAGAG GAGCCCAAGA 780  
GAGACACTGA GGCAGAGAGA TCACACAGA GTACATGCTC CTGGCTCTGT TTTCAAGTGT 840  
GCTTTGACCA GGAATATATG ATTAATATAC TCCATPACG GTTTTCCGAT ACAAAGTGC 900  
TAGAATAATAC ACAAAATTC CCAATGCTA AGTTGTGCTA ATGTCTTTCC AGTTCTGGG 960  
TTGGGAGTG GAGGTGCGA GGGTTGTTT GTGCGCAAC GTTCAGTCTC GTTTCACAGG 1020  
AGGAATTTGA GTCTCCAGG GTCTCATAT CCGAGTGAT TGTACAGGGA GCGCTCTGCC 1080  
CTGTCTGCTC TCAAGTCCAG GAGAGCTTTG AAGCAGTCAA GCTTGTGCTT TGTACCCCAT 1140  
GTGTCTGCTC TTTGTTGAT CACTCAGAG TCACCTCTGG ACCTCTGGGG TTGAGTTTCC 1200  
AGTATGCTCT TATGGGGGCC CACTCACT GGTGGGCTGA GTGGAGCTC CTTACCATG 1260  
TCCCAAGAGA CACTGAGGTG CTGGCTTTT TATGTGCTC GTTTGTGCC GTAGTTCTTT 1320  
TGTAGTGTCT CATTTTGGCA TTTGGCAAT CAGCCTGGAA GTCTGGGCCC ATGACAGCAA 1380

1440 TCACTCTCTC CCAAGCTCC TGAAGCTAGA GGAAGATTG CTGAGATCCA TTATTAAG 1440  
1500 CAGGAATGG TGTGCAATG AGCTGCATGG TTTAGGAGT CTTTGGAGC CTTCAGATC 1500  
1529 CTCAGAGCA ACAAATCTTG TACTAGAA 1529

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 245:

20 GTGCGAGTTC CCCCAGGCC CCGAGCGGCC TTCCGCGCCC GGGGGCTCC CAGAGCAAC 60  
GAGGCGCTG AGAGCTCCAC CTAGTTTACA GATTAATATC CCAAGAGACA ACTCGAGTTC 120  
ACCAATGCT AGGCCCCAGG TGTGTTAGC TCTGTATTA ATGTCTAGC TGTCTGTAAA 180  
TGGCCCTGAA TTTTACCTTT CAGGTTATTC TTCCAGTTAC ACAAAATCT ATGAGATGG 240  
TTGTGAGAT TATCTTACTC TATCGAATA TGTTCAGGAT TTTTGAATC ATCTTACAA 300  
GCAAGCTGCC AGTTTGTAAA CTGAATTTCA ACAGTTTCCA GAGACCTTCA ATGCTGTGTT 360  
TACACAGAT GATGCTTTGC AGAGCTTGT GBACTCATC TATCGAAGG CCACTCAT 420  
CCCAATTTC TCTTATATGG GAGCTGCTCT GTGTAAATAC CTGTCCCATC ATCTGAGAT 480  
TAGCCAGAG AGTGGCACT TCGCCATTT CTTACTTTCA AGATGTGGA CTGAATATGA 540  
AGTTAAAGAT CAGCTGCGAA AAGGGGATGA AGTTACTGGA AAGCGATTTT ATGCAATTTT 600  
ACTGTTTCTG GCGAGCTTTT ATCTTACCT GAGATCAGG GBAAGAAATG GAGAGGTTAC 660  
AAGAGAGAT ATTCTTCAGG TTGCTTTGG AGAATTCCT AATGCCCTGT TTCTAATCC 720  
TATGATGAC AATTTAATTT GTGCAATTA ATGTTTAAG TTGACAGAT CAGTTTGA 780  
AGATGCTTGG AAGGAAAGAG GAAGATGA TATGGAGAA ATATTTCGA GAATGTAAA 840  
CTTTGTCTTA GATCGAATCT GCAATAGAGA TGTAAAGAG ATGCTCTTCA AGCTTGTAGA 900  
50 ACTGCGTCA AGTAAGCTGG CCAAGTCCA TCCAATCTCA ACATATAGAG AAGCAAGCC 960  
AGAAATGAT CCTTACTACT TTATGATGA ACCAATATTT TATACATCTG ATGCTGTCC 1020  
TTTCACTGCA GCTCATCCAG ATTACAGGA GAATTAACCA GAATTTACTT AAGAGAGGA 1080  
CTTTTGTCCA GATTTATGAG AAATGTGAC AGATTATTC GGGGCTGGTG ATCCATATTT 1140  
GATATATTT GATGATGAGA TGGAGCCGA GATAGAGAA GCTTATGAAA AGTTTGTGTT 1200  
60 GGAATCAGG CTTAAGCGAA AAGATTAAG TTAATTTCA GATATCAT TTTTAAGAG 1260

AGTTTAGSTA TGGTGAATTA GCGAGACACA AGAGACGACG AAGATGTGTC ACATCTATAC 1320  
CAATTTAGG AGTTGAGTT ATGTACTATA TGTATGCAAC TTATATTTTG TTATACACTA 1380  
TCTGCCAAA TAACTTTAT TCCCTATAC TTAAATGTCG TAAATATATA TAAATAGTTTA 1440  
TTATGTACAG TTATATCTAC TGTATTGGCT GCATTAATAT CAATTTTGA AATAAATPAA 1500  
AAAAAAAA AAGGAGGCC GCTGTAGAGG AAGCCAG 1537

15 (2) INFORMATION FOR SEQ ID NO: 246:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 506 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 246:

25 TCGAGGATTT GGCAGAGACC CCGCGGGTGG GCGGTTGCTA TCGCTTTGCA GAACCTACTC 60  
AGCGAGCAG CTGAGAGAG TTGAGGGAAA GTGCTGTGCG TCGGTTCTGCA GAGCGATGG 120  
ATAGGTGCA GCGGAAATTA AATCATGCC CTTTCTGCTT CAGTGTGAAA GGCACAGTCA 180  
AGATGCTGG GCTGATATT ATCACTATC TGTATACAC AGTATCTATG CTATCTGTAT 240  
CTGTGTTGG ACTGATACA GAACACACA CATGTACAGT TGGTGGAGGG GTGTTTGCAC 300  
TTGTGAGGC AGTATGCTGT CTTGCGAGG GCGGCTTAT TTACCGGAG CTTCTGTCCA 360  
ATCCAGCGG TCGTTACAG AAAAAAGCTG TGCATGAAA AATGAGATT TTGTATTTT 420  
ATATTACTTT TTAGTTTGAT ACTATGATTT AACTATTTT CTGATTTAT CCAAAAAA 480  
AAAAAAAA AAAAAAATTT TGGTGG 506

45 (2) INFORMATION FOR SEQ ID NO: 247:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1348 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 247:

55 GTCTTCTTTT TACGTGTTTG AGTTGTGAG TGAATGATA GGGTAACATG GGGCTTCAGG 60  
ATGACCCCTT GGAAGTGTGC CAGATGCTTT AATCTGAGC TGGATCTCG GACTGTGGAG 120  
GCCCCGTGTA GGGGAGCTTC TGGAAAAACC TGGGAGTTGA TCCCGAGAGT TGGAGAAC 180

TCCTCTCAG GGCAGGCTGC CTTGAGACAC TGGTAGTTCT GGGGCTGGGA GGGAGAGGG 240  
CTCGGCTTTT CTCTGAATAG AAGCTGCTC TTGACAGTTT CAACTACTTG TTCTCAAAAC 300  
ATTTCTAAT TGAATGAGAG GTTTTCATTA GCATGTTTTC TTTAAGGCAT GGAAGGGAA 360  
GAATGCTCAA GCAATGCTAG TTTGTTTCCA GTGGATGGG CCGGCTTCT CACTGTGCGG 420  
GGCTTCCCTT TGCATGTGGC ACCTTTGTC AGGGCACCA GGCAGACTCT TCCACCTTTC 480  
TCCACTGAA GCACCAAGGG GCTTGAGCGT TAATTTGGCT AATCAGAGC ATTTTTTTGG 540  
TCTAGTATC TTTCAACTT GTCAACCTT CTATTTTITT TAAAGTTCT GTTCTCTGTA 600  
TTAACAGAA ACTAGAGAGA AATGTTTCT GAGGCAAGTT TATTTGTAG ATCCCAAGG 660  
GGAGTTGCG TACAGAAAA TAGTAGCTG GTTTAGAAAC TGACGAGGCG AAGACGCCAG 720  
GAGCATGCG AGAGGAATTT GCCAAGATC TACCTTGAGA TACGCTCTGT CCAATGTCTT 780  
CACCACTGA ATACCAAGG CTCGAAGTG TTTTCTCTCT TTGAAAAA AATTTCCCA 840  
AGCTTTTAAA GGTGCATTTA AGATCATG TGACTTTAGA ATGGAAGTGC GGGCTGTGC 900  
AAGTGTGAG TGTCTAGAA GTTTCAATC CTCTGGAATG CATGTGATAC TCAATCTCAT 960  
TTTGTCTCT TGAATGCAAT TTTGTTTCT TACGAGATCT GTCCCTGTGG GTGCTGTCTA 1020  
AGATGTGGA CACTGTGTT TTTGTGTAG ATTGAGCTGG GAGCTGCA TCAAGTCTTT 1080  
TATATGAAA TTAGGACCA CCAATCTGT GTCCCTGCT TGTGTGCTAA TGAATGTAGG 1140  
GAGGAGGG ATGTCAAGCC AATAGTAGC CTTCCCATG GCTTTGGCCA GGCAGAGAC 1200  
TTCACTGTT TTACATGTT CTGTGTAAT TTAAAGTTTA TGTGTATATA GCGAGCTGT 1260  
TTCTGTGAAA CTGTATATT TGTAAATAAA TATATTGCTA CTTTGAGAGT AAAAAAAA 1320  
AAAAACTGCA GGGGGGCCCC GTACCCCA 1348

45 (2) INFORMATION FOR SEQ ID NO: 248:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1766 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 248:

55 GTCCGAAATC GCGAGAGCG CACGAGCGGC CACGAGACA GCGGGAATTA AGGCACTTCA 60  
GCGAGCAAT TCCCGAATTA TTCTATTTCC CTTCTCTCTC TCCCGCCCCG TATCTCTTTT 120  
CACCTTCTC CACCTCTGC TCGGTATSCA TCGCGAGCG TCGCGGACCA CTCATGTCCA 180

5 TTCCATCTCC TCGTCTGCT TCGAGCCGA GCGCTGCGG CCGCGCGCG GCGGAGGCC  
AGAGCCCTCG CCGCGCGCG GAGCGAGAG CTGACGCTCC TCGTCTGCGG TCGACGATCT  
GATTCTCTCG AGAGATGCG AGAGATGCG GTTCTGCTTT GCGACGCGC TCGATGCT  
GCTTCTCGG GAGCTTTCA GTGTCTGAG TGTGTTTCT TACCTGATCC TGGCTCTCT  
CTCTCTGAC ATGAGCTTCA GGAATGACA GTGCTGATC CAGCTGTAC AGATGCGA  
10 AGAGGCGAT CATTCTGAG CTTACTGGA GTTAGAGAT ACTCTGTCT CAGAGCTTT  
CGATTAATC ATGATGCTG CGATGCTCA CATTCTGAG GCGCTGATC TGTATTTG  
15 TCTCTCTG GTAGAGATC TGTGATGCT GTTAGAGCT GCTCTCTCA TGTGCTGAT  
GAGCTATGTT GTGCTGCTTT TTAGCGAT CAGCTTCTA ATTCTTCTG ACTCTCTAT  
70 TTCTAGTCT CGATGCTCT ATGAGAGTA CAGAGCGAG ATTGATCAT ATTCTGCTAT  
CGCGCGAGT CAGACGAT CATTCTGGA AAGATGCA GCGAAGCTCT CTGATGCT  
80 CAGAGAGG GAGATGAG TACTGTGAA CAGAGATCT AAGATGCT AAGATGCT  
25 TTAATGTTA TACGCTGTT ACTGTGCTA TCGAGGAAA TACTGATGCT CAGCTGCG  
CTGATGCTA AGCTTTTCT TTAATGCT GTTCTCTCC ATCTTTTCC TTTAGCTCT  
100 AGATGCGC ACGAATGTT ATGAGCTAT AAGAGCTA TCTTAGACT CAGAGAGTA  
1080 AAGATGAA TTTAGAGT AAGTCAATC GTTAATGCT GTAGAGCTTT TACCTGAG  
1140 TCGAGGCG AAGATGTA GTTAGAGG AAGATGAG AAGATGAG AAGATGAG  
1200 TCGATGTT AGCTGATC TTAGCTGCT ATGATGTA GTTCTGCTT TACGATGTA  
35 TCGAGCTTT CAGCTGCT TTTCTGCT CAGAGCTG CAGCTGCT CAGCTGCT  
40 TTTCTTTAA TCAATTAAT ATCACTGTA TAAATTAAT ATAGTGATG GTGATTAAT  
CTGATGCTA GATGCTAT TGTATTTAA GATTAATAT GGAAGTGGG ACTGAGGG  
45 AGTACGCG CATGCTGCG TCGCTGCT TCGCTGCT ACTATGCT CAGAGGAA  
1560 RCGTCCGCA TTTGCGAG TGTGTTTAC GTCACTGAG ACCGTTCTG AGCTTAAT  
1620 TCGAGACTG TCGCGAGT TGTGTTTCT CTTCTGCT CCGCTGCTA AGTTATTA  
1680 ATAGGTTCT ACTTTCTTA CTATTAATA AAAAAAATA AACTGAGGG GCGCGCGTA  
1740 CCGAATGCG CGATTAAT CTAA  
1766

(2) INFORMATION FOR SEQ ID NO: 249:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2664 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 249:

5 AGTGTCTCG GAGAGCGG AATTAAGGA CTGAGGAG CAGATTCGG GATTATCTA 60  
10 TTTTCCCTCC CTCTCTCGG CCGCTATCT CTCTTACCG TTCTCCGAC CTGCTCGCG 120  
TASCTGCGG GAGCTGCG GCGCTGAG TCCATTTCA TCTCTCTCT CTCTTCTGA 180  
GCGAGCTCT CCGCGCGCG CCGCGCGCG AGCGAGAG CTGCGCGCG CTGCGCGAG 240  
AAGAGCTGA GCTCTCTCG TCGCTGAG GATCTGATTT TCTGAGAGA TGTAGAGAG 300  
ACTGCTTTC TCTTTCGAC CAGCTGATC ATGCTCTTT CCGTGCAGC TTTGATGTC 360  
20 ATCACTGCG TTTCTACTT CATTCTGCT CTCTCTCTG TCGACATGAG CTTCAGATC 420  
TACAGTCCG TCACTGAG TGTACAGAG TCGAGAGAG GCGATCTAT CAGAGCTTAC 480  
CTGAGCTAG AATTAATCT GTCTGAGA GCTTTCATA ATTACATGA TCGTCCGAT 540  
GTGATCTA ACGGCGCTT GAAATCTAT ATGCTCTCT TCTGCTGAG AGATCTGTT 600  
GACTCTTA AGCTGCTCT CTCTCTGCG CTGATGCTT ATGCTGCTCT TCTTTTAA 660  
30 GGAATGCG TTTCTATCT TCGTGAATC GTCAATTTCA GTCTCCGAT TGTCTATGAG 720  
AATTAAGTA CCGATTTGA TCAATGTT GCGATGCGG GAGTCTGAG CAGTCAAT 780  
35 GTTGAAGTA TCGAGGAA ACTCCCTGA ATGCGAAA AAGAGGAGA ATAGTATCAT 840  
GGAAGCTGA AATGAGAG TTTATTAAC ACGATTAAT AGTTATTAAG TCGTTACTG 900  
TACTATGAG GAAATTAAT AGTCTGCT TCGCTGCA TCCAGCTTT TTTTCTTAT 960  
40 TTTGCTTTT CTCCGATCT TCGCTTAA CCGTCAAT CAGGAGAA AATTAATGA 1020  
CTGATTAAG AACTATCTTA GACTGAGA GAGGAGAA TCAATTCAT AGATTAATC 1080  
45 AATACCTTA TGTGCTGAG GCTTTACT GTAGCTTGA AGGAGAGAA TTTGAGTTAA 1140  
GAGGAAAT GAGGAGAG CTCTGCTCC TTTCTGCGG TTTTCAAGC TAGTCTTCT 1200  
CAGCTATCA TTTATGTTT GCGCTTGA AGTCAATAT AACTATTA AATTAATTT 1260  
50 GCGAGAGA GTGTGATC TTTCTGCTT TTTTCTGCA GCGCTCAAT CCACTCTCT 1320  
TCCCGAGA TGTGAGAG TACCTGAT ACTCTCTTT TTTATGAT TAACTATCA 1380  
CTTGATTAAT AACTATGAG TGTATGAT AATCTGAT TCGAGAGT CCACTGATA 1440  
AAGAGATA GAAATGAA GTGAGTGA GAGGAGTA GAGGAGTCT TCGCTGCGG 1500  
GTGATCTCT CTGCGATAT CCGAGGAA GAAAGCTC GCGCTTTG GAGAGTCTT 1560  
60 TCTAGTCA TCGAGCGG TTTGAGAT TATTTGAGA ACTCTTTCC GATTTCTTT 1620

5 TCTGCTCTT CCGTGGCCA CTTCAAGTTT AATTAAATAG GTTGTACTTT TCTTACTATA 1680  
AATTAAATGT CTGTACTGCT TGTGCACTGCT TGTAAACTTG TTAGAGAAAA AATTAACTCTG 1740  
CATGTGGGCT CTTCAAGTTT TGAATTTTTG TGAUCCATCT TCAUATCTGG GGGGAACTAT 1800  
CTCAAGAGGT GAATTAACAG AAGCTTTTTT TCTTTGATCT TTTCGCCAGA TTCAAAATCTC 1860  
CGATTCOCAT TTGGGGCCA GTTTTTTCT TCACTCTCAA TATGAGATTT CAGCGAATTT 1920  
GAAGAAAA TCAATCTGTA GTTCTTTGAG GTTCTCACTC ATATGTCATGA TCTTTTCAGAG 1980  
GGAAATATCA CTGGGAGTT TAAATTAAGG GCTATGATAT TTGATGCTCC CAAGTAAAGG 2040  
CAGCTCCAAA AAGTATGAGA AGGAATTTCT CTAGCTGTCT TGGAAAAATT AGTTAAGGAT 2100  
TTGATGGGTT AAAGATAC CTTGCTTTAC TCAATCTTAT TTCTTTAGCC CCGTTTGAT 2160  
GTTTTACTG GTTTCATGTC CTATAGAAA GTGCATCTC CATCTCATC CTCTGCGCTC 2220  
CGAGAACTC AGTATGTC TTTTGGGCT TCGCTCCAA AGGACTTCT GCAATGGAG 2280  
TGCACATCC AGTCTTTCT TTTTGTGCT GCTGTGTTTA GATTAATGA GAGATCTTTG 2340  
TCCACACAG GATTTTTTTT TTTTAAAG AAGACTATA GATGAAAT TACTAATGAA 2400  
ACTGTGTTA CGTGTCTG CTGTGACAT AATATACAG TACGACTTAA GAGCTTGAA 2460  
TCTGTGTTCC TGTAAATTTT CAATTCATG TGTATTTAT TAAAAAAA AAAAAAATA 2520  
AAAAAAA AAAAGGCGG CCGCTCAGA GATTCAGC TTAGTAAAG GTGCATGCA 2580  
CGTCAATGAC TCTTTCTATA GGGTCCGCC AATATCATT CAGCGGCG TCGTTTTTAT 2640  
AAGGTGCTG ANTGGGAAA ANCC 2664

## (2) INFORMATION FOR SEQ ID NO: 250:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 865 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 250:

50 CCGTGGAGT AGTAAACAGA TTACGCCAT TTGGCCCGA CCGCTCTTT CTGGAAATCC 60  
GGGTGCTGCG GATTAAGAT CCGTTCTTA ACGGTGGAT CCGTGTCTC GGGATGAGAT 120  
TTGGGCTTTC CTGCGGCTT TGGTGGATC GGTGCTCA GATGAGAT TTAGGCTTTC 180  
TCGGGGCTTT CCGGATCTC ACTTAATAT CCGACTCCA GATGAGAA GCGCGGAACC 240  
TAGAGGCTT GATTAAATG GTGCTTAC TGGTCTGTC AGTGTCTCG GCGATGCAA 300

5 TGTGTGAC CTGTCTCA GGTGTCTG TTTTGGAG CTTTCCCA CAATCTTG 360  
GACTAGTCA GAGAACTC TTCCCTTCT ACTTCACAT CTGATGGG TGTGCTTCA 420  
TCAATCTG CATGTGCT TCAAGCAT GTTGGGTCA GCTCAATTC TGGAGGCA 480  
GCCAGCTT CTTGTTTC CTGAGCTTA CCGTGGCAC TGTAAAGCC CGTGTGCTG 540  
AAGCCGAC CAGAGTCC ATGTGGGCC TCAAAACCT GAAGAGAG CAGAGCTG 600  
GTGGAGGT ACGAGGAC CACCAAGT CCGATCTCA CCGCAGCTC CAGAGAGG 660  
ACCCGAGTA CAGTCTCT CCGCAATTT TCTTCCCTA CAGTGGCTG TCTCTCTTT 720  
GCATCTGG CTGCTCTG AGCAATGGC TGTGTCTG TGGCTTGG CTGAAATTA 780  
GAGGCTCTA GATGGGCC TCAATCTTA TAAATCTTC TTCAAAAA AAAAAAAA 840  
AAGCTGAG GGGCTCGGT ACCCA 865

## (2) INFORMATION FOR SEQ ID NO: 251:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2082 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 251:

35 TGGGGGGG AATGGTCT TGGCTGAG ATTCGCAAT CTGAAATTC TCCATACTT 60  
GCTACTGT TTTTTTTT TTTTATCA CCGCCCGCC CACCCCGCG ACTTGCAGA 120  
TGTCAAGA TCTGACGA GTTCTCAT TGAAGCTG ATCACTTT AAGCTTCAT 180  
CATTCATA TTTTCTTC TCTTCCCT TCAATCAT AACTGTCT CATTTCTGT 240  
CTGTGTGT GTTTTATTT GTTGGATT TTTTATTA TTTTACTTT AGACTTGT 300  
GTGTGCCA CTTTCTTC AAGCTCAC CTGCTCTT CTAAACCT CTCTTCCAG 360  
ATCAAGAAA AAAAAAGA AGTTTTTT TCTTCTCT GATTTCTCA TGTGAGAT 420  
AGCTGAAA GAAAAAAA ATGTGAAT TTAAGACTT GCAGCTGCC GAGTTCATC 480  
GGTTTTTT TTAGCATG TTATCTTA ATAGAGAAA AATGCTCAT GAACCTTCA 540  
CAATAGCC TCAATCAC TTTCTGCT GACTGTGAG TTTTGGCTT GTATGCCAA 600  
ATCTAGAGT TTAGCTGCC ATTAAGAAA CTCAATCTA TCTCATCAT TATTATGCTT 660  
GCTACTGT CTTAAGCA ATGACTATA ACTTTTCTA AGACTTTATG GAAGAGAG 720  
ATTATTTA TAAAAAAA AAGCTTCAT GCTGAGCT TATGATTA TTTTCTTC 780  
CTTTTTTT CTTTGGCT TGGAAATGA CTTTCAAG CTTATAGAT GCAATCTA 840

5 CTTTGTGTTT ATTGCTCAAT GACTTTTGTG AGTTTAGAAC AAAACAGTGC ACCGTAGAG 900  
CCTTCTGCC ATGAAATTTT GCATCTGCTC CAAACTGCT TTGAGTTACT CAGAACTTCA 960  
ACCTCCCAAT GCATGAGAG CATCTCTGT GCAAAATAC CAGATGGGT TACAGTTTA 1020  
ACCTGGCAA CATTGAGAA CTCCTTAAGT TTCTTTTATA ATAGAAATCA GCGCCCACTT 1080  
10 TGGGCACTTA ATTGTGCTA TTGCGAGAA CGAGCTTAAA ATTATTTT TAAAAAGAA 1140  
AACTGCCCC TTATTTTGG TTGTTTAT TTATTTTATA TATTTTGTG CTTTGTGCA 1200  
TTGTGCAATG TGGATGCTC TGGGTTTCTA GTATATAT TTATTTCAAT TTATTAATC 1260  
TTTAGGCCA GTTAAATAT ATGCTACAGA TAAAGGAATG TTATAGATA ATTGAAAGA 1320  
15 GTTAGGCTG TTATGCTGTA GATTTTATA AGATGATG CACTAAATG TTATGATTTG 1380  
TGATGTTAAG GGGGTAGAG TTGCAAGG GACTGTTTAA AAAAGTAC TTATACAGA 1440  
TGCTTTGCA ACTTAATAT AGTTGGGTA TGTGTAGTCT TTGCTATACC ACTGACTGTA 1500  
20 TTGAACCA AGATATTAAG AGGGAAAG CCGCTGTTA TATCTGTAGG GGTATTTTAC 1560  
ATTGAAAT GTATGTTTTT TTTTCTTTT AAATTAAG TATTTGGAC TGATTTGAC 1620  
TAAGATATA CCGCAACA TATATACA AAAAATTT CAAACTGTT TAGAAGCTA 1680  
30 ATAAATTTA TCGATTTA AAATGGAT TACTGACAG TTATAGATG ATGCAATTT 1740  
TTTACAGTT GTATGTTGT GCAAGCTG ATTTCTGTA ACTTAAAAA AAATCCGAG 1800  
TTTAAAGC ATATATAGT AATGTTAT TTCAAGGACT GACATCTGT CTTTAAAG 1860  
35 AAATGAAG TAAATCTAC CAAATTAAT ATAAATAT CTTGTCAATT ACTTTCTTT 1920  
TACATATTT GCTGTGAAA ATGTGTTAT ATCTTGAGT ACTACTAAC CACCGTGT 1980  
40 GTTCTATGT GCTTTCTTT CATTTCAT TCTGTTATA TCAAGAAAG AATAATCTAC 2040  
AATAATAAC GCAATTTTT TTGAAAAA AAAAAAAAA AA 2082

## (2) INFORMATION FOR SEQ ID NO: 252:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1482 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 252:

50 CAGCAGGCT GCGCCCGGG ACTCTCTCT GCGCTCTCT CTTGAGGG CTTGCGGCTT 60  
GCGCCCTCG CCGCTAGGA GTCTTAGCC AGGATGAGG CTTGTGTGA CTTGTACAA 120

5 GAGGTGATGA AGCAGGAGAA TCCCGGATC CAGGGCTACC CTCTGATGG GTCCGCTCTG 180  
CTAATGACT CAACTCTCT GACTAGCTG TACTTCTGTC TCTGCTGTC GCTTCGATC 240  
ATGGCTAATC GGAAGGCTT CAGCTCTCT GCTTCTGTA TTGTCTACA CTCTGATCTG 300  
GTGCACTCT CCGCTCAAT TGTCTATGAG TTCTGATGT GCGGCTGCT GAGGACTAT 360  
10 ACTTGGGCT GTGACCTGT GCACTATTC AACAGCCCTG AGGCACTTAG GATGTATCG 420  
GTGCGCTGCG TCTTCTCTT CTTCAAGTTC ATTGAGCTGA TGGACATGT GATCTTATTT 480  
CTCGAAGA AAGAGGGCA GTTGACTTC CTACATCTCT TCAATCACT TGTGCTTCC 540  
15 TCGAGCTGT GTTGGGGGT AAGATTTGCC CCGGAGGAA TGGCTCTTT CAACTCGATG 600  
ATAACTCTT CCGTCAATGT CATTAATAT CTTTACTAG GATTAATCT CTTTGGGCT 660  
GTGCACTAC CTTACTTTG GTGGAAGAG CACTGACAG CAATCACT GATCTCATTT 720  
20 GTCTGCTCT CACTGCAAT CTTCCAGTAC TACTTTATGT CCACTGTAA CTACCAATAC 780  
CAAGTCATA TTCACTCAT CTGATGTAT GCACTCTCT TCTCTACT GTTCTCGAC 840  
25 TTCTGTATC ACTCTATAC CAAGGCAAG CCGCTGCTCC GTGCACTCA CCAAAATGGA 900  
GCTTCAGGTA TTGCAAGGT CAGGCGCAC TGAGAGCCT GCTCTAGATA GCGGCGCAC 960  
TAAGTCTCT AGGACTGAC CTTAGGGCAG TGTCTCTCAG TCGCTCTCT ACCTACACT 1020  
30 GTGACCAAG CTATGTTGT CAGGACTGAG CAGGCGACTG CCGCTCTCT CCGCAGACT 1080  
GCTTACAGG GACACAGCT TTGTTCTCT ACCCACTTC CCGGCGCAC TCGAGGATG 1140  
35 TCGCTCATT GCTGTCTCC ACTCCAGAG TGGGGCTTA AGGCTCTTA CAGTTATTTT 1200  
CGCTCTCTG CTTAAACT TGGGAGGAA GCACTCAGG CTGGCGCAC AAGGCTCTC 1260  
GTGCGCTTTT TCTCAGACA GAGAGCTCA GCAATATGT CACTGTGAC CAGTCTCAC 1320  
40 TCTCCAGCC CACAGCTGA AGCAGTACT TCTGGGCGCA AGGCTCAGGT GGGCGGGGC 1380  
CTGGAAATAC AGCTGTGGA GCTGCTTAC TCACTCTGT TCTTATTTA AGTGAACGA 1440  
45 GGAAGCCAA AAAAAAAAA AAAAACTGA GGGGCGCGG TA 1482

## (2) INFORMATION FOR SEQ ID NO: 253:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 834 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 253:

50 GGCAGAGCG CCGTTCGCG CCGTGGGCTT ACGAGTCTT TAGCCAGAT GAGAGGCTTT 60



5 GTGAAGTCTT ACCAGAGGT GATGAGGAC GCGAGATCCC GATTCAGGG CTACCTCTTG 120  
ATGGGGTCCC CCTTGTAT GAGCTCCANT CTCTGACT ACCTGTACTT CTTCTCTCA 180  
CTTGGGGCTC GATCAGTGG FATTGCGAG CCCCTTCAGC TCGTGGCTT CATGATCTC 240  
TACAACTCTT CACTGTGGC ACTCTGGCTT TACATCTCT ATGAGTTCTT GATCTGGGC 300  
10 TGGCTGAGCA CTTATACCT GCGCTGTGAC CTTGAGACT GCACCTTAG GCGATGTTCG 360  
TCAGTGGCTT CTTCCATAC ACTGTGACC AAGGCTTAGT TGTTCAGAC TCGAGAGGG 420  
15 ACTGGGGCTT CCGTCCGAC AGCTGTCTTA CAGGGACAC GCGTTTGGTT CTTGACCCAC 480  
TTGCGCGGGG CAGCTCCAGG GATGTGGCT CATTCCTCT TCCACTTCA GAGCTGGGG 540  
CTAAGAGGGT TGTACAGTTA TTTCGGCTT CTTGGCTTAA AACTTGGAG AGGAGCATC 600  
20 AGGGCTGGC CCGACAGGG TCTGTGGC TTTTCTCTCA CACGAGAG GTGAGCATTA 660  
ATGTCTACTT GAGCCAGTC TCACTCTCT ACCCCACACA CTGAGCAT AGCTTCTGG 720  
25 CCAAGGTCA GGTGGGGG GGGCTGGGA ATACAGCTTG TGGAGCTGC TTACTCACT 780  
TGTCTCTTAA TTAAAGTGA CAGAGAAC CACGAAAAA AAAAAAAAA AAAA 834

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1508 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 254:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 254:

TTGAAGTCTT AAATTTTAT ATCCGAAAC TCTAGATCC TAGATGGAA GCTGTCTCT 60  
ATTTCTCAT GCTACCTTC CAGGTCCAG GAGATGGTGA AGAGCTGCA CCGCGAACA 120  
45 CCACCAAGCT TGGAGTGA CCTCTCAT GAGCTTCTTG AGACTTTTG CAGATCTCC 180  
AAGTGGTCTG GTCCGAGC CTTTCTCTT GTCTGCGAG CTGTCTTGA GGTAGTCTC 240  
CTTCCCATGG ACGATTTC TGTCTCTCT ATGCGCATC ATGCGCATC TCTTACTCTT ACGAATGAC 300  
50 AAGGTTCTTA AGTGGAGT CTTCTCTCA AAGACATTA GACAACTCT ACTAGAAAA 360  
GACTATTTCT TGGCTCTG CAGTCTGAC CAGAGGCTG TGGAGCAGC CATCTGGCT 420  
55 CTTGAGATGG ACCGTGAG GATGTGAG TATTTTCAA GCATCCACC TCGCATTAC 480  
AAATTTCTG AAGTCCAT GAGCAGAG TCTCATAGT ACTAGAGG TGTGATCTG 540  
60 GTGTCTTTC TGTCTCAT AGAGCGAGG TTAGTGGG ATTGTGCG CAGTGTACT 600

GGATAGCTT TCGCGAGG AGAGACTTC CTCTCTGG GACTTCATG CAGGTGCGAG 660  
TTCTCTACAC CATTACAG GATTTACAG ATCGAAGGA AGTACAGTA AACACTATTA 720  
5 TCTATCTTG ACTTAAAGG GAATTAATTT CTGAGAGAT TATATTTGTC ACCGAGCCT 780  
TAATCTCTTC TGTCTCTG ACTGATGAA ACTTGATTT GAGAGCATT TTCTTATGG 840  
10 AAGGATGAG ATTCCGAGG ACCTGCAATG CTTTCTCTG GTTTTATTTA ACAAATGACA 900  
AATGAAATTC TTACAGCTG AAGGAGAGG TGTGCGCAGA TGTGAAGAG ACCTTCACTA 960  
TCAGCCCTAA CTCTCTCTC CCGAGAGA CTGTGTGGC TCTGTGGCA CCTGTGAGC 1020  
15 CCGAGCTCT GTTGAAATG TTGTGACAT GTCCAAATGG GAAGCAGGG GTTTTACAT 1080  
CTCTTAAGG ACCTGATCC AACACAGTA GATTTGACTT AAACCTCTTA GCGCAGATA 1140  
TGTCTTACA CATTACAGA ATGTGTCTG AGTGTCTGT TCTGATTTT TCACTCTGCT 1200  
20 CATTGACTGA AGGAATTTA TTAGAGTAT ATGTATCTC TGTGTCTTTT AACTGTATCA 1260  
TATCTAGCTC TCGAGTCAA CTTCTTACA TACTGTACAT ACCTGTGAC ACCTGTGGA 1320  
25 GTCTGCACT CTTTATCAT CTTGTTTAA CTTGTGTGG AAGATTTCT TTGTCCAAAT 1380  
AAATTTAT ATTAGATCT ATAGAGAGG ATATATACAC TTTTGAATG TTCTTAGATG 1440  
TCTAACATA AATGCATTT GTGACTCTA TTAATAAAA TTAATAAAA TCGAGGGGG 1500  
30 CCGGTAC 1508

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2514 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 255:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 255:

GAGAGACTCA CACTTCTTTT CATTATCAC TACCATGTA GTGAGATG CAGGGGAGA 60  
GCACCTACT GTTGTGTGA GATTGTGTA TGAATCTCAT AACCTAAGAG AGGAATTTAT 120  
50 AGGCTTCTG CTTATGAG CCGATCGCA AATTTGGCT GTGAATTTT ACATATGAT 180  
AACTGAGAG TGGGATTA ATATGAGTA TTGTGTGGC CAGGCTTACA TTGCTCTAG 240  
TGTATTTCT TCGAATGA AGTTGTGTC TTCTAGACTT TTAGAGAT ATCCCGAAGC 300  
55 TATCTACACA CTCTCTCTT CTTGTGCTT AAATATCTG TTGCAAAAT CAGTACTCT 360  
TATGAGATA TGTGTGAT TAGGAACAT TAGGAAGTT TTTTCTTTT TCACTGATC 420  
60 ACCCAACTG CTTTATGAC TTGACAACT AATTTCTGT CTTTTCAGA ACATTAAGA 480

5 AUGGGGTAAA GAGCTGAGG AAGTCTGCA TTCTCAGTGG AGAGGAGGG ATGAGCTTT 540  
TGAATTTTAA GTGAACTCC TCGAGCACT TGTTTTATGT TTAGATGCTA TAAATAGTGA 600  
CAGCAATATT AGATGGATA ACTATAGCC TGGCCGAGCA TTCTGACTCT CGAGTGCAGT 660  
GTCAAGTTT GTTTCTGTT TTACTATGTT TGTCTTAAA AATGCTCAT CTTTTACAG 720  
10 AGCCTTTGGG AAAAAGCTCC AGGGGCAAC CTCGTATGTC TTCTTTGCGG CCGGTAGCTT 780  
GACTCGATTA CTCGATTGAC TCAACCAAGT GATTGGAAA TATTGAGTT TATCTGAGT 840  
TTTGGTTTGA GAGGCCAACA AATTGGCA GAACTTTGA TATTCMAATG AATCTCCCTG 900  
GGAAATTCGG CAGAGCTCAC CAGGCTTACT TCGAATCTCA CTTAAGCTCT GAGCTTACT 960  
ATTAAGGAAC CTAAAGTGTG CCAAGAGTGG AGCAGATTAT TCAAGAACTT AAGATATAT 1020  
20 TCTCAGAAC GCACCTGAAA GCTCTTAAT GCTTATCTCT GTTACCTCTA GTCATGGGAC 1080  
AATCGAATT CATTAGCTGG GAGGAGCACC ATGCTGAGAT GTATAGAAAT GACTTACCCA 1140  
ATCTGACAC GCTGTCACT GAGCTTCATT GTTGAGGAT CAATGGAAA CACAGGGGGA 1200  
25 AAGATATGA GCTTGGCTCC ACATCTATG AAGCCCTCA CTTGCTGAC ATCGAATTTT 1260  
TTCTTAATGT GTATGCAATG CTGAGCTCC TGTATATCT TCTGTGATG AAGTTGAGA 1320  
ATGAGCGGTA TGAATATGA GAAAGCTTC TTAAAGCATA TTGAGGACAC ACTTTGAGG 1380  
ACCAAGGTC AAGTAACTG GCTTTGCTTA ACATAAATT TGAATATAA CACGAGCTGG 1440  
30 ATTATATGT GAGCACTAT ATTAAGCTCT ATACAGTTAA GTCAAGCTT CTCAGATA 1500  
ATTCCGAAC TGTGGAAAT ACTTAAAGA CTTTAAAAA TAGGCTTCT TATATTGAT 1560  
ATTGGAAGA AAAGCGGTA AGCTATGTA GACCACTTAA TCACATAATA TCTTTGCTTA 1620  
40 TAGGACTCCA TTGAATACAT TAGCCATTGA TAACTACTCT GTTAAATGG CCGCTGTTG 1680  
AAGCTCGAG CTTTGAAGAC CTACCTGTT TTCAGAGGA GAACTTTGAA AGTGCATGT 1740  
45 TTCTTTTTC GTATCTGTG TTGATGGAC TCTGGAAATG TTTCAATTAA GTCAATTTAG 1800  
ACATAGCNT TATATCACT GTGATCTCT ACTGTGCGG TGTATGAT TCTTTGAGA 1860  
AATATATTT GAAGAGGTT GAGAGGAGG AATACATTTT ATAAATGTT GTATGAGAC 1920  
50 CCAATATTA CTTTGACTA ATAGAGTTT TAAATGTT TAAATCTAT ACTGAGAT 1980  
TACAGAAAT TACCGGAAA AAGCTTGTGA GCTACCAAA CAGGATTC AGTGATNT 2040  
TTCTCTTCT TGAATTAA GAAGCAATG ACAAAGTTG AATGGAAAG CCTGCTGTTG 2100  
55 TTCCACAT CTGTCTGTT TACATGCT TTGTGAGCT ACATCTCT AGCTTTTAA 2160  
GCAAGTAT GTTGAOACT TCTGTTTCA GTTTGAGCA GAATGAGG CCAATGATAC 2220  
60 TGACAAGTA TTGTCTGTT TTTTCTCT GTCTTTTTC ATGACTCTTA TATATCTCT 2280

5 CATCTGATT TATAGCAAA ACTCGAAA CCTACAAAT AAGTGTGTG GTTATCTAG 2340  
AAGATATGG AAGATATTC TGTATTTT GTTGAAGAA ATCAATTTG TATATTTAT 2400  
TTCAATCTAA ATAAATATG AATTGTGTT AAGCTTAGG CACATTAAT TTGTGCGGT 2460  
10 CAAACNTTC TTGTATTAAT TCTTTAATC AATTGATAA CAGTTCACA ATTC 2514

(2) INFORMATION FOR SEQ ID NO: 256:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2157 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(+1) SEQUENCE DESCRIPTION: SEQ ID NO: 256:

CTGCTTAATG AAGCGATGC AGAATTTTG GCTGTGAAT TTCACTAT GATPACTGAG 60  
25 AAGTGGGAT TAAATATGA GTATGTGCT GGCAGGCTT AATTTCTC TAGTGCAATT 120  
TCTTCAAAA TGAAGTTGT TGTCTTGA CTTTATAGA ATATGCCA AGCTATCTAC 180  
ACATCTGCT CTTCCTGTC CTTAAATAT TGTGTGGCA AATCACTACC TTTATATGA 240  
30 GTATCTGTC CATTAGAAC AATTGAGAA GTTTGTCTT TTTTCCATG ATCAACACA 300  
CTGCTTTAG AACTTGACA GTTAATTTCT GTTCTTTTC AATACATTA AAGAAAGGCT 360  
AAGAACTGA AGCAATCTG CAAATCTAG TGGACAGCA GCGATGTC TTTTGAATTT 420  
TTATGAGAC TCTGCAAGC ACTGTGTTA TTTTATATG GTATTAATAG TGAACAAAT 480  
ATTAGATGA ATAACTATAT AGCTGGCCA GCATTTGTAC TCTGCAATGC AGTGTCAAT 540  
40 TTTGATTTCA TTCTTACTAT TGTGTGTTT AAAAATGTC TATCTTTTAC AAGAGCTTT 600  
GGGAAAGCC TCCAGGGCA AAGCTGTAT GTCTCTTTC GCGCGGTAG CTTGACTGA 660  
45 GTACTGANT CACTCAACA AATGATGGA AATATATGA GTTTATCATG AATTTGCTT 720  
TGAAGAGCC ACAATTTTG CAACAACT TGAATTTCA ATGAACTCC CTGGAAATTT 780  
50 CCGAGAGCT CACAGAGGTA ACTTGAATC TGAAGTACC TGTGAGATTT ACTTAAGA 840  
AAGCTTAAT GTTCCACAG TGAAGACAT TATTCAGGA CTTAAGATA TATTTCTGA 900  
ACAGAGCTC AAGGCTCTTA AATGTTATC TCTGTACCC TGAATCATG GACACTGAA 960  
55 ATTCAATAG TCGAGAGAC ACATGCTGA CATGTATGA AGTGAATAC CCAATCTGA 1020  
CAGCTGTCA GTTGAAGTTC AATTGTGAG AATCAATG AAGACAGGG GGAAGATAT 1080  
60 AAGGTTGCG TCCACANTCT ATGAGGCTT CCACTGCTT GACATCAAT TTTTCTCTTA 1140

1200 TCTGTATGCA TTGCTGAGG TCTGTGTAT TCTTCTGTG ATGAGGTG AGATGAGCG  
1260 GTATGAAAT GCACGAAAG GTCTTAAGG ATATTGAGG AACATTTTGA CAGACGAAAG  
1320 GTCAATAC TTGCTTTTG TTAACTAAT TTTTGTATTA AAACAGAC TGGATTAT  
1380 GGTGACACA TATATTAAAC TCTATACAG TAAGTCAGAG CTTCCTACAG ATATTTGCGA  
1440 AACTGTGGA AATAGCTAAG AGACTTTTA AATAGCTT TCTTATATTT GATATTGGA  
1500 AGAAAGCC GTATGTATAT GTAGACCACT TAATCACTAA ATATCTTTG CTATAGCACT  
1560 CCATTCGATA CATTAGCCAT TGTAACTTA CTCTTTTAA TGGCCCTGT TTGACCTTC  
1620 AAGCTTTGA GACTACCTG TTCTTCCAG AGAGAGCTT GAAAGTCCA TTTTTCCTTT  
1680 TCCCTGACT CTGTGATGG CACTCTGGA TTGTTTCACT TAAGTCATTT TAAGCATAGC  
1740 ATTATATAT ACTGTGACT TCTACTTGT GCGTGTATG AATTCCTTGA AGAAATATAT  
1800 TTTCAGAGG TGTGAGGA AGCAATACAT TTTATTAAT GTTCTAGTGA AGCCCAAT  
1860 TGACTTTGA CTATAGAGG TTTTAAGTAT GTTAATAAT TATACTGAC AGTTAAGA  
1920 AATTACCGA GAAAGCTG TGAGCTCAC AAACAAGAT TTCAATGAG ATTTCCTTT  
1980 TCTTGACTT AAGAACAA ATGCAAAAT TTGAATGGA AAGCTGCTG TTGTTCACA  
2040 TCTGTTCT GTTATATC CTTTTGAG CTATATCTT CTTAGCTTT TTAGAGTGA  
2100 TAAGTTGAC ACTCTGTTT CATGTTGAG ACAGATCAG AGCCATGGA TACTGACAC  
2160 TCAATGACT GTTATTTTC TGTGCTTTT TCCATGACT TTATATAGT CTTATCTG  
2220 ATTATAGC AAGACTGGA AAACCTGGA AATAGTAT GTGTTTATC TATGAAATTA  
2280 TCGAAATAT TCTGTATTT TTGCTGAG AAATCAAT TTGTATATTT TATTTCAATC  
2340 TAATTAAT GTATTTTG TTAAAGCTT AGGCACATTA TTTTGTGCG GGTCAAMCA  
2357 TTCTGTGTA AATCTC

(2) INFORMATION FOR SEQ ID NO: 257:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 689 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 257:

60 ACTTCTGTT GCAAAAGAT GTTCAAGCT TATTTATAC TTGCTGCCC CTTTCCTTT  
120 CATTTATGG ATGAGCTGC AGCTCTAGA AGACTGTTT TTTGATAG AGATAGCAT  
180 CAGGAACCG GATGTGGTG CAGGCTGTC TCTGCTGT TCGAGATTG TCGACCCGG

240 AGCTCTTAGT GGAAGAGCT AAGAGATATG TCCAGTACT TCCATCTCT TCTCTGTCTC  
300 CGATTTTAGC CCACACACC AGGTAGCTT CCAATTTTTC TCTCTTTCA TAGCTGTAG  
360 GCGTTTCTG GGAATGTTTC TCAATCTCT TAATCTATTA TTGGTCAAT TTTCCTCAT  
420 GTCCCAAGC TCCATCACT GGCACCACT CCCACAGAG ATGCCCTCT CATTCAGCTG  
480 GCGCTTGAC TCCCACTG TGTACCTCT TTGTGTGAC GCGCTGCTC CAAAACCTTC  
540 AGCAACAGC TTTCAAATG GAATTTTCA CTGTGAGGS CTTTACATC AGCAACGCA  
600 AATCTCAAT GCTGCTGAG GTCTGCTCT ATTAAATGC AATAAATATG TAAGTACATA  
660 AAACAGCAA TAGAGAAC GTATGCTTT ATTCTCAAT ATGATATCTT ACATAGAAA  
689 GCGAAATTA TTAAAGATAG TAAGCAATT

(2) INFORMATION FOR SEQ ID NO: 258:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2377 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 258:

60 TCGACCCAG COTGCGGA TGTATATTT CTGCTATTT CCAAGAACG GCGCTATGC  
120 ATCTCTTCA TATCTTCAAC TGTATAGG GAGGCCCGG GCGCTGTCT ATCTGTGCG  
180 GCGCAAGCTT GGTGTGTTT TGTCTGTTA CTGTGCGG CCGTCAAGCT GCGTGTCTC  
240 TCTTGAAGA AGCTGTATC TATCAAGCT GCTGACAGC ATCATCTACA GTCAATTCG  
300 GCGCTACCTG ATCAATCTC TCCAGCTTC GCTGTTCGG AGCGCGTGG GAGCCCGCT  
360 GCTTTGAAG TCTATCTC CATGTGGG GAGGAGGAG CTTTCCCTCA GCGATTTGG  
420 GTGAGCCCC AGAATCTCT CAGGTGCTT CAGAGGTTC AGCTGACAG CTCCACAGA  
480 CAGGCCATA TCGAGAGCT GGTTCCTAT GCGATGTTT TCTCTGAGC TGAAGATTT  
540 CAGAGCTCT TCAAGAGCT TGAAGAGCT GTGTTAAG AGCAACCGC GAGGCCGAG  
600 TACAGCTCT GATTCTCA GAGCGACCA GTTCTCTTC GCGCACTNAC TACTTTGACT  
660 AGCTGGGGA CTTCAAGC CTGCAAGC TGTGTGAT TTGCTGTTC CTGCTCTGG  
720 ATCGAGATG TGTGCTGTC TGAAGTGT GACTTCATCC TGGGGGTCT CAAGCTGCTC  
780 TTGATGTGT ACTAGCTTT GAGATCTG GCTCAAGTC TTTTGCCTG GCGCTTCCA  
840 RGATACTTCT CTTAACCCA CAAAGTGT TTGAACGGC TCTCAAGCT TTGTCTGCG

5 TGGAGGAGG GATCTCACT CTGGCTGTG AGCGATGCC AGACCCAGGC TGGAGGCCG 900  
AAATGGTGGG CTGTGTGTG CTGTGGACA TGAACCCGAT ACTGAACATG CTCAATCTGT 960  
TCCGCTTCTT GGTATATAT CCGAGATGA AGCGATGGC GTTGTGTGGC AATACGCTTC 1020  
TGGGCTGTGT GCAAAACATG CTGTGCTTGG GCGGATCTCT GTTGTGTGGT TACTATCTAT 1080  
TTGCAATCAT TGGATCAAC TTGTATTAGG GGTTCATTTT GGTCTTCTCT GGAACAGACA 1140  
GCTTGGCCCC TGGCAATAGG TGGGCGCTCT GTGGAGCTTT CAGGAGCTG GATTACTGGG 1200  
CCAGACATTT CGATGACTTT GGGGCTGGCC TGTTCATCTT GTGGACTTGT ATGTGTGTGA 1260  
15 AGACTGTGCA GTTGTCTCTG GATCATATAT GGGCTACTTA AGGCGCTTGG TCCAGATCT 1320  
ATTGTGTATT GTTGTGGTGT GTTGTCTCTG TCAATCTGGT CAACTCTTTT CTGGCGCTGA 1380  
TTCTGGAGAA CTTCCTTCAAC AATGGGAGCC CCGGCAAGCA CTTCAGGCC CTTCGTGGA 1440  
CCGCAAGGC CACTTACAG ATGACTGTGG AGCTCTCTTT CAGGATATTT CTGGAGAGC 1500  
CGGGAGGA TGAAGTCA CAAGAGCTGA GCGAGAGCC GACCTCTTGG CTGTGAGGT 1560  
25 GAGTTCGGG TCTGCTTCC CAGAGGGGC GCGAGAGAG AGAGGCTGG ATACAGAGG 1620  
TGGCATAT GAGAGAGGG GCGATCTGT GCGAGAGAG GAGAGAGAG AGCTTCTCTC 1680  
TGAGGAGCA CTAGCTGGG GACAGGAGCC AGTCTCTTGG GTTGTGGCC AACAGCATT 1740  
TACAGAGAG CTGTGTGTG TTCAAGGAGG GCGGTGGCC TCCGCTTCTT TTTATAGCTG 1800  
CTTCAATGAG AATTCCTTGG TGGACTTCA AGGAGCTTT CAGACAAA TCCAGAGAGC 1860  
35 AGGCGCTCC CTGTGCTCT GAGCTTCTG TGTGTCTTT GTTGGCCGCA GCGCTTGGG 1920  
ACGACAGGCC TGAACAGGC CTGACAGGT TACGCTGAG ACTTCCGGG CATTCAGCTG 1980  
GAAATGATAC TAATACCTCC GATTTAGCC CAGCAGACA GGTATAGTTC CAGTTTATT 2040  
40 TTCTTCCAT AGCTGTAGG CCGTTCTGG GATGTTTAT CATTCCTTT ATCTATATT 2100  
TGGTCAATTT TTCTGTGATG TCCGAGCCT GCGTCACTG CAGCCACTC CCGCAGAGA 2160  
45 TCCGCTGCTC ATCGACTGG GCTTTGACT CCGCAGCTGT GTACCCCTCT TGTGTGAGG 2220  
CGCTGTGCC AAACCTTCA GAAAGAGCT TTCCAAATGG AGTTGTCTAC TGTGAGGCC 2280  
TTTACATCA GCAAGAGCA AATCTAGTG CTGTGAGGG TCGTGGCTCA TTAGATGCA 2340  
50 ATATATATGT AATGATATTA AAAAAAAAAA AAAAAA 2377

(2) INFORMATION FOR SEQ ID NO: 259:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1193 base pairs  
(B) TYPE: nucleic acid

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 259:

5 TCTGATGCC GTGCGCCGCG CCGTGGCTTT TGGCGGTGG GCGGAGCTT CTTGTGTCT 60  
ATTTCAGGG ACTCCAAAGC GAGCGCGGGG ACTGAAGGTG TGGTGTGCA GCGCTCTGGC 120  
10 AGAGGTTTAA CTTGGGTCAA ATGCAAGGAT TCTCACTGG TACGATTTAG CTGTCCCGCG 180  
GCACTGCCG GAGGATTTGA AGTCTCTGAG CTTCAAGTTT GATCTTAGTC CAGAGAGGCG 240  
CATGAGGTG CCGCAAGCGG CAGCGCGGAG CTTTCTCTGT AGGCAATTT GCTATTTTCC 300  
15 CCGAGTCTTT GCTGCCAGAG CTGTGACTGC GAAATCGAAA GTCTTTCAAG AGCTTCAGAA 360  
GCGCTTCCG TACCTCCGCG AGCCTATTTA CCGGATCTT GATGGAGCC GCGTCCCGGA 420  
20 GCTTTTGGC AAGAGACAG TGAACACTAG TCTGAATGTA TACCGAAATA AAGATGCTTT 480  
AAGCAATTT GTATTTGCG GAGCTGTGAC GCGAGTCTTT TTTAGCATTA AGGTAGGCGT 540  
GCGTGGTGG TGGTGTGGG CATTAATGGA GCTTGTCTGG GCACTCTCTT AGAGGCGCTG 600  
25 CTATGGCAT TTCAAGACTA CTCTGTGTGAG ACTGTCTGG AAGAAACA GAAGATGCA 660  
AAGGACTGC ATGAGCTTAA ACTGGAGAG TGAAGGCA GACTACAGT TACTAGACAC 720  
30 CTGCTGGA AATTTGAAG TTGTTTAGAG GAGATGAGC CTGAGATGA TCTTAGAAA 780  
ATTGAGCAC TGTAAAGCT TCTTAGAAC CTTCACTTAA TAAATAACA AGACAGGAC 840  
TGAATGCT CTGAATTA ACTCACTGG AGAGCTGAG GAGCTGCCA TGTCCAGTGA 900  
35 ATGCCAGAG AGAGGCACT CTTTGTGAG CTTCTGACA AATTAGTG CTGCTAGCTG 960  
TGTGGCACT GCGTGTCTCT TGTCTTTTTC TTTCTTTT TACTAGAT GCGGCTGTGG 1020  
40 TACTCTCACT TTACTATCC TTAAATTTAA ATCACTACTT ATGTTTGTAT TAATCTATCA 1080  
ATATATGCAT ACATGATAT ATCCAGCCAC CTAGTTTTA ACCTTAAT AAACATTTTC 1140  
45 GCAAAAGATT AAGTTGAT TTTACATTA AAAAAAAAAA AAAAAA AAA 1193

(2) INFORMATION FOR SEQ ID NO: 260:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1262 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 260:

GAAGAACTCA AAGATGAGA CAATCTCTTT GACATGAT TGGGGGCTCT CAATATGCT 60

GCATTACTAC GAAAGAGAGA AAGGCAAGT CTCTTAGTA ATCTTGCC ATGTCTTAG 120  
GGTTTGTCT TGAACGGGA TTCTGCANT GAAAGCAGC TTGTTAAMA TGAAGAGCC 180  
5 ACCATAAAC AAGCTTACAC GAGTSTCCA ATGTAGACA ATGAATTACT TGTATTAGT 240  
CTTGCTTAT TTAAAGGGA GACTACTGCG CATCTCCAG GACATGAAA GACTGAGAT 300  
ATAAACTTT CAGCTCCAG TATCCACAG GACTGTGTG TGTCTTAGA CCGAGTTCA 360  
10 ATATGTTAT TTGTACTG TCTCTTCCA GAGTGATA TTCTTTTGA AAGTCTTTG 420  
GTTTGACAG CATTAAGTGC AAGAGGAA AAGTTTATC ACCGCTGA AAGAGTGA 480  
15 GAATTTGAT CTTTAGAGC ACTAGTTTG GCACTTAA GATTTTACT TATTTTATC 540  
ATAGTATTG AACTCATGC AATAGATGT GAAACACT AGATTAGTA GTTTATCTG 600  
GCGTTTGT TAAACTGA GATTTGGA AATGTTGTC ACTGCTTTC CAGCTATGA 660  
20 ATATTTTGT GAATGAGC CAGGATTTA TGTGTGATC ATCCATPAC ACGACAT 720  
TTATTTAAA AAGATGTG TATCAAGT AATGCTCAG ATGTGAGT ACTATGTTT 780  
25 AAGACCAAG TGAAGGGA TGTGCTGA GTTGCGTGG TATGTTTAA GCGAATTC 840  
AAGAGAGA AGCAAAATA GTTTTGTCT TTGAAGAT TTTTAAAT TATTTATG 900  
GTCTTTTTT TATTAATAT GTGTGATG TTACATGTA TGTGAGAT CTTTGACCC 960  
30 TAAATGCTT TTGTGTATC AAGATGTG TACTATTTT ATTTTATA ATGTATCTT 1020  
CCCTTTCTT GTTTAGAT TACTTCTC TTGTTATC TTAATCTGA TGAATGAGA 1080  
35 CATTAAGT CATATAGA TGTTCATG TTCAATAT TACTCTTG TGTCTATTT 1140  
GTTGACAGC TTAAACAGA GTTGATGTA CTTCATAT ACTCATGA TACTTAGGG 1200  
CAGTTCTTT GCGATGTGG CTTTGGAA GCAAAAAT TACCCUAG GCAATCCCA 1260  
GT 1262

## (2) INFORMATION FOR SEQ ID NO: 261:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1179 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 261:

GCAAACTTT CCCCAGGC TTGAACTT GCAAGCGAA ACTTGAATC GTTAAAGTT 60  
GGTTGCAAC GCGCCCTCG CCGAGAGG GCAATTTGC GTTCCGGA GTTGGCCCT 120  
55 CAAAGCTCG GAGCGAGCC ATGTCTGCA CCGAGACAG CCGCCCTCG CTAAGAGAC 180  
60

CTGACCTCA TCTTCTCG CCGACTCG CCGGAGGG GAGTTTACA CTGTGAGA 240  
CGTCTCTG GACTGCTGT TGAATCTTT ACCCGAGGG GTACAGAG AGAATATC 300  
5 ACGACTCAG CTCAGGAG CTATATGCA GAATATGTT AAGTGTGCA ATGACTCTG 360  
CGATGAGT CTATATCC TGTCAACA GATGCGAA ATGTGAGC TGAATTTT 420  
10 GATTTCCCT CCGAGCAGT TTGAATGAG TGTAGTTCT TTGAAATCA AATTAGACT 480  
TCTTCTCT TTTTATGAT GTTCAGAA CCAATGACT GAGCATTT ACCCAGAT 540  
AATCGGAG AGCTCTATG CGAATTCCA GGAAGCTTT GATCAGCTT GTACAGAT 600  
15 CATTCGAC AGGACCCAG AGGAATCG AGCGAGGC CTCTTACT ACTCGACT 660  
CTTGTGAG GCTTTAGGC CCGCTCTGA TGAATCAG ACCCTTCAA GTATATGT 720  
20 TTTCAGTT TCTATGACT TCTCAGAT TGAAGCAG CAGAGAAC TGAATCTTA 780  
TTTGAGAC CACTTTGTG GATTCGAA CCGAATAT GATATCTCA TGACCTTCA 840  
TGAATGTA AATGAGCA CAGTTGTCT GATGAGAT GAAAGAGAC AGCTTTAA 900  
25 CTTATCAC ATCTTGCTA TCGGCTGT AGCTACCA AATGCAAT CTATGTGC 960  
TATGTACT TCTATTAC AGCAGCCC CTATGACA GATGCCACT TTGAAATTA 1020  
30 CTACATGCA CAGTTGAG CAGTATCAC GTGCGAA CAGACTACT CACTGTCT 1080  
ACCTGCAAT TAAATATCT TTAATATGT CTTGTGGA AGCAATTT CAGAGACAG 1140  
35 GAGAGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA 1179

## (2) INFORMATION FOR SEQ ID NO: 262:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1162 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 262:

GCAAACTTT CCCCAGGC TTGAACTT GCAAGCGAA ACTTGAATC GTTAAAGTT 60  
GGTTGCAAC GCGCCCTCG CCGAGAGG GCAATTTGC GTTCCGGA GTTGGCCCT 120  
55 CAAAGCTCG GAGCGAGCC ATGTCTGCA CCGAGACAG CCGCCCTCG CTAAGAGAC 180  
60 CTGACCTCA TCTTCTCG CCGACTCG CCGGAGGG GAGTTTACA CTGTGAGA 240  
CGTCTCTG GACTGCTGT TGAATCTTT ACCCGAGGG GTACAGAG AGAATATC 300  
ACGACTCAG CTCAGGAG CTATATGCA GAATATGTT AAGTGTGCA ATGACTCTG 360

CCGATGAGGT CTTATATGCC TGTCAACAA CAGTGGCAGAA AATGTGGAC TGAATTTGT 420  
GGATTCCTTC CCGAGGCGAT TTGATATGAG TGTAGATTTCT TTTCAGATCA AATTAGACTC 480  
TCTTCTGCTC TTTTATGAT GTTCAGAGAA CCGATGACT GAGACATTTTC ACCCCAGCAT 540  
AATCGGGGAG AGGCTCTATG GCGATTTCCA GGAAGGCTTTT GATCAGCTTT GTACAGAGAT 600  
CAATTCCACG AGGAACCCAG AGGAATCCG AGGGGGAGGC CTGCTTAAAT ACTGCCACT 660  
CTTGTGAGG GCGTTTAGGC CCGCTCTGTA TGAATTCAG ACCCTTCAG GGTATATG 720  
TTCCAGGTTT TTTATGCACT TCTCAGCAT TGGAGAGCG CAGAGAAAC TGGATCTTA 780  
TTTTCAGAAC CACTTTGCG GATTTGAGAA CCGCAAGTAT GATATCTCA TGAACCTTCA 840  
TGGATGCTTA AATGAGAGCA CAGTGTGCT GATGGGACAT GAAGAGAC AGACTTTAA 900  
CTTATACAC AATGCGCTTA TCGGGGTGT AGCTGACCA AATGTCATC CTATATGCG 960  
TAATGTCACT TGTATTTACC AGCAGGCCC CTATGTACCA GATGCCACT TTAGCATTA 1020  
CTACATGCA CAGTTTCAG CAGTATTCAC GTTCCAGCAA CAGACTTACT CACTTTGCT 1080  
ACCTGCCAT TAAAGATCAT TTAAATATGT CCGTGGGGA AGCCATTCA GACAGACAG 1140  
GAGAGAAA NANGAAGAG AG 1162

(2) INFORMATION FOR SEQ ID NO: 263:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 735 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 263:

CGGGCTGGGT ATTGCTCCG CACCATGGG CCGAGGGCA AAGTGGGAC GAGAGGGAG 60  
AAGCAGATAT TTGAGAGAA CAGAGAGACT CTGAAGTTCT ACTTGGGAT CATACTGGG 120  
GCCATGCCA TTATCTGCT TGTGAGTTG GTCTCTCTTT ACTCATCTG CTCATTTTG 180  
GCGTGGTGG CTTTGGGCTT TAGTCTGCA GTATATGGG CAGCTTACA CTCATTCAG 240  
TGGATGGAC GAGCAGGCTT CTTCTGAGA TGGGGCCCTG ATGGATGGT GCGAGGCTC 300  
AAGCTGAGG AGGGCATGCG AGAGACTCT AAGGATGTGA TCTTACTAC AGCATCTG 360  
CAGGTGCTCA GCTGCTCTC TCTATGTC TGTGCTTCT GGTTCCTGCG TCGAGGCGG 420  
GCCCTTACC TCGTGTGGT GAATGTGCT GGGCCCTGGT TCACTCAGA CAGTGGCAC 480  
CCAGACCGAG AGCAGATGA GAAGCGAG CCGGAGCAG AGCGGCGCA GATGAGGCG 540  
TTATAGCAAT TGAATTTGT GGCAGAGCC ACTGGCCCTG GGTGGCTGT TCAAGGTGA 600

CAGCCCTCA TCCCTGAGC AATGAGGTC TAGTCCAGG GCCAAGACA GTCTGAGTA 660  
TTGGTATAC TTATCTCTA TAGGTGCTT GATTAATGG CTTGATATGT GAAAAAAA 720  
AAAAAAA ATTTT 735

(2) INFORMATION FOR SEQ ID NO: 264:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 783 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 264:

AAGTGCATGA GGTCCGATG TGTGCTTAG TGAATGGGT TTGCTGGCT CTCCTGTGT 60  
TCCCGGCTG GGTATTTGC TGCACCATG GCGCCCAAG GCCAATGGG CACGAGAGG 120  
AAGAGCGAG TATTTGAGA GAGAGAGAG ACTCTGAAT TCTACTGCG GATCATACT 180  
GGGCGCAAT CCAATTACT CTTGTGAGG TTGTCTTCT TTTACTCAT TGCCTCTTT 240  
TGGGCTGGT TGGCTGGG TTTATCTCG CAGTGTATG GGCAGCTAC CACTCTATGA 300  
GCTCATGGC AGAGCAGG TTCTCTGAG ATGGGGGCT GATGATGAT GCGATGGAC 360  
TCAACATGA GAGGGCATG GAGAGTAG TGTCCCCAC CCGCAGCCA GCGACCTTAA 420  
GATGTGATC CTACTGAG CCAATGTGA GTGTCTGCG TGTCTCTCT TCTATCTCTG 480  
GTCTCTCTG CTTCTGGCT CAGGCCGGG CCTTTACTC CTGTGGTGA ATGTGCTGG 540  
CCCTGCTTC ACTCCAGCA GTGGCACCC AGCAGCAGAG CACATATGA AACGGCAGG 600  
CCGAGAGAG CCGGGCGAG TGAAGCGGT ATAGCCATTG AGCATTTGC SAGTGGCAC 660  
TGGCCCTGG TGGCTCTCT AGGTGCACA GCGCCCTCAT GCTGAGCA TGAAGGCTTA 720  
GTTCAGGGC GAAAGCAAT CTGAGTATT GGTATTAAT ATACTCTATA GGTCTCTTGA 780  
ATA 783

(2) INFORMATION FOR SEQ ID NO: 265:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1638 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 265:

5 GGCACAGGCG GCGCGCAGCG GTGCGCGCGG CCGCCCCCGG CCGGAGCCCTT NCCCTTTTCC 60  
GTCTGGGAGC GCGCGGTGCG GTCTCAGCGG ANCCCGCGGC ACGGAGAGCG GGAAGAGAT 120  
GAAATCCCGG GCGCTCCCGC CGGATGCGAA GAGGAGGAA GTGATCCGAA ATCTGCGGCT 180  
AGTCTCGC AGAGCGATG TCTACTACTT CAGTCCAGT GTTAGAGT TCAGAGCGAA 240  
10 GCGTCATGTT GCGAGTACC TGGGAATAC TGTGATCTC AGCAGTTTG ACTTCAGAC 300  
TGGAGAGTGT ATGCTTAGTA ATTATAGAA GACAAACAG AGACTCGAA AGATCTCT 360  
15 CAATCAAAAT AGCGTAAAC CAGCTTGA TACAACTTG CCAATAGAC AAGAGGATC 420  
AATTTTCAAA CAACCGTAA CCAAGTCAAC AATCACTCT AGTAATAAG TGAATACAA 480  
CCAGAGCA ATCAATCAAC AGCCAGTCA GCTTTCTCG GAGAGAGCG TCAAGAGAT 540  
20 TAGTCAATCA GATGTACAG AACAAATAT AAAAAAGT GAACTACCA AGGTCTTCA 600  
AGCAGTCTGT CAGGTAGCA ATGATGAC CTTTATCT CTGTCTTCA GTCTTTTCA 660  
25 CAAGCTCT GCGCAATCA CAGGCAAGT CTGCGCTCT GTGAAAAG AGCCTGCT 720  
TTGGCTTAC ACATCTAAC CCGTCTGCA AGCTTTAT AGCTTTAT GTCLGNT AGACATCAG 780  
840 GAAAGAGAA GAGCGATAC AGCAATAG CAGAAATG GAGAGACAC TGAATGAGA 840  
30 CAGCTCTGCG CGAGCTGT ATGAGAGA GATGATAT GAAATGAA GTGAGTCA 900  
AGCTTAGAA TATGATAG TAACTTTGCA CCGACTTTC CAGAGAGAA ATTCTAGAA 960  
ATTGAGAAA AATTTTCCA CTGCTTTTG CCGTTAGAA AAAAATGTA CCGAGGACA 1020  
35 TAGAGCTTT TAAATGACT AACAAATCC TTTTATGAT TATTTTGTAT GTATATATCT 1080  
ATTATGAAA AATCATGTT TATTTGAGT CTAGAGCTT AATATAGTC TTTTATATA 1140  
40 TCAAGAGAA CCGTAGATG AGCTGAGCT TTTGATGCA GTGCAATCT ACTGAAATG 1200  
TAGCACTTAC GTAAAGATT TGTGTCGCC AGATTTTAA TAGAGACGA TCAAGATTC 1260  
45 TAAATAAAT TCCGATTTA AGATTAATGT GACTTCACT TATATAAAC TATTTTATA 1320  
CTTTATGAA AGGGACAC TGTACTTCT TCCATCTCA CTGTAAAGC AATATATCA 1380  
TATATTTAC AGATGATG GATTTCTTC TGTGAAAG CAGCAATAT AAGAGACCC 1440  
50 TGTATGCT TCTCTGAT TACTTCAAC TGTATGCG GCGCTTAG TTTGACATG 1500  
GAGTGGGAG GAAGTAGCG CATATATG CAGTATGAC TATTCCTCT GCGAGTTCT 1560  
GAGGATCT GCTTCAACA GAATTTTCA GATTTCTG CTTAATATAC ACTTAGGCTG 1620  
TGTATATTT TTTTCCCT 1638

60

## (2) INFORMATION FOR SEQ ID NO: 266:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1455 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 266:

5 CTTGCTGACT GCGATGCGG TACCGGTCC GGAATTCGA GGTTCGACC ACCGTCTCC 60  
TCAGTTGCGA AGGTACCTCG GAAATACTGT TGAATCTAGC AGTTTGTACT TCAGAACTCG 120  
15 AAGATGATG CTGATTAAT TACAGAGAA CAAACAGAGA CTGCGAAGCG ATCTCTTCA 180  
TAAATATAG GTTAACCCAG ACTTGATATC AACATTTCCA ATTAGACAAA CAGCATCAT 240  
TTTCAGACAA CCGGTACCA AGTCAAAA TCATCTTAGT AATAAATGCA AATCAGACC 300  
20 ACAAGATG ATGACAGCC CAGTCACT TTTCTGGAG AAGAGCTAC AAGCACTTAC 360  
TGCATGAT GTACAGAC AATATATA AACCATGGA CTACCGAAG GTCTTCAGG 420  
25 AGTTGTCCA GTAGCAATG ATGAGACCT TTTATCTCT GTTGCAGTG CTTTGCACAC 480  
AAGCTCTCG CCATCAGAG GCGAGTCT CCGTCTCTG GAAAGAGCC CTCTCTTTG 540  
30 GCTTAACCA TCTAACCC TCTCAAGC TTTTATTTCT ACAGATGAG ACATCAGAA 600  
ACAGAGAG CAGTACAG AGTACGCA GAATTTGGA GAGCACTGA TCGCAGCAT 660  
CTTGTGCGA GCTGCTGTA CAGAGAGAT GATATTTGAA ATGAGCAGTG CAGATGAGC 720  
35 CTAGCATAT GATCAGTAA CTTTGGAGCG ACTTTCCCA AGAGAAATT CTTAGAAAT 780  
GACAAAAAT GTTTCAGTG GCTTTTGGCT GTAGAAAAA AATGTACCC GAGCAGATG 840  
40 AGCTTTTAA TAGCACTAC CAATGCTTT TTAGATGAT TTTTATGTA TATATCTAT 900  
ATTCAAAAA TCAATTTAT TTTGATGCT AGCACTTAA ATTAGTCTT TGTATATCA 960  
AGCAGACC TAAAGTGA CTGAGCTTT GATCCAGGT GCAATCTACT GGAATGTAG 1020  
45 CACTTACGTA AAGATTTCT TTTCCCGACA GTTTTATTA GAGCAGATCA GGAATTTCTA 1080  
ATAAATTTT CAGTTAGCA TTTTGTGAC TTTACTGAT ATTAACATAT TTTTATATT 1140  
TATTTGAAG GAGCACTCT ACHTCTTCC ATCTCACTG TAAAGACAA TAAATATTA 1200  
TATTCAGCA CTGATGGA TTTCTTCTGT TGAAGAGC ACAGATTA GAGCCCTCG 1260  
TTAGCTTCC TCTGATTTAC ATTCAACT GTTCCGCGG CTTTATGTT GACATGAG 1320  
55 GTGCGAGAA GATAGCCAT ATATTTGCG TATGATCAT TGCCTCTGG ACCTTGTGAG 1380  
GAATGTCT TTAACGAA TTTCTAGGA TTTCTGCTT AATATACG TACCTGTG 1440  
TAAATTTTT TCCCT 1455

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 268:

5 GCACCGGAG CAGCGGGCT GGTCTCTCG CAGCGGGCG GCGCGGAGT GCGCGGGCA 60  
GGAACAGCA ACCTTGGAGT TGGCAGCAT GAAATGAAAT CAAATACCGG TGTCAATAC  
ACCGGGGTA TGTGGCTGAC ATATGCAATG GAGTTGGCT TGGTCTATAT TGTCTTACTC 120  
AGCAATCCCT TCTTCAGTGT TCTGTGTCT TGGACTTTAA CAAATATAT ACATATATCT 240  
GGATGTACG TATTTTTCGA TCGATGAAA GAAACAGCTT TCGAACTCC TGAACAGGCT 300  
AAAGCAGAG CTCCTTAACTC ATTGGAGCA ACTGGACTAT GAAATACAT TTACATCTTC 360  
ACGGAAGTTT TTCAATATTT CTCGAATAT TCTATATTTT CTGGCAAGTT TCTATACGAA 420  
GTATATGCA ACTCACTTCA TCTTAAGAC AGCTTCTCTC CTGAGTGTAC TAATTCGCA 480  
AATGCAACA CTACATGCTG TTGGATCTT TCGAATTTAT AGTATTTGAA ATGTTTTGA 540  
ACTGAAAAA AATTTTACAG CTACTGAAT TCTTATAGG AGGAGTGT TAGTAAACTG 600  
CACTGTCTCT CTGATATGT GAATCAGAA GTATTTTACAT TGGAGGCGCA ATGGCTGCTC 660  
CTTCAGTGC TGTTTTGAAG TCCAGATTC CATTAATGTA TGGCTCTCTT TAATACACT 720  
GTATCATTC TGAAGAGGG CTTTATAGC AGGCTGGCGA GCGCGAGCTT ATAGTTTAA 780  
GCGCTGACA GTGAGGCTGT ATGATGAAA TTCAAGGAAA TAAGGATTT GTAGAAACT 840  
AGGAGGCTT TAACTTATA TGAATGGCA TTGTGTTAG AAGAGCAT TTCAAGTCAAT 900  
TCAGCTGTGG TTATTTAAG CAGACTTACA TGTAAACCGG AATCTCTCT ATACAGTTT 960  
ATTAAAGATT ATTTTATTA CCGTAAGAA AAAAAAAAA AAA 1003

(2) INFORMATION FOR SEQ ID NO: 268:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1214 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 269:

ATCAGACTCT ACAATGACA TATTTTGAT GTGTGTGTG TGGTACTTCA AAGTAACTAG 60  
GAAAAATTA TCTTCGCAAC ACAGTACCT TGTCAATGCA GAAATGCGGG TGTAAAGTTG 120  
CCAGTTGAT CAGTGTGAT TCAATTCAT ACTTCTTACA GAGCAAACT GAACTTTGCA 180  
GTTGCCACA CTGAATGGA TCGAATTAC CTTGTGATGA ACAGCGCGGG TATGTGCTGT 240  
ACATATGAT TGGAGTTGG CTGGCTCAT ATGTCTTAC TACGATATCC CTTCTTCAAT 300

(2) INFORMATION FOR SEQ ID NO: 267:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1086 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 267:

5 GCGCTCAGT ACCGTCCGG AATCCGGGG TCGACCGAG GGTGCTGAC CCGAGGAG 60  
CTGCTGTCT ACATCAGCT GGGCTGGAGC GCGCTGGGCG GCACTGAC 120  
TATGTCTCT TCAAGGGCGG CAGCTGTGT CATTCATCT TTATACCCCA GCACTGACA 180  
GTGTGTGAT GTATGAGCA CTGAAGAA GTGTGTGAA TGAAGATCC CTGTGAGAG 240  
CAAGCGACT TTATCTTTC CTGACCTTG CTCTATGAC ACACCTCTC CTGACTGCA 300  
CTGTACTCC TTGAGAGAG AACTCTCTA GGGAACTTG ATGGAGACA GCGATGGCA 360  
AGGACTCTT GGGTGAAGCA GGGTACACT TTGATCACT GAGCAGCTG AGGCTGTGG 420  
ACCGAGGT TACGAGAG ACATGAGC TGAAGGAGA GTGAAAGAC TTGTGTGACA 480  
AAATTTGCCA GTTTCAGAA ATAGTGGTG GTTATTTGA GCTGTGTAT CAACTTCAA 540  
AAGAGCGA AATGAAAG ATGAAGGCA TGGTGTCTG GACTTGTCT AATCTATAG 600  
CAAGCAGAG AGAAGCTCA CAGCGCAC TTCAAGCCT AATGAGAGA AAGAAATCC 660  
AGCTGAAG GTATCGGTT GATATGAG CTTTGTGAA ATGAGAGCA GACAAAGTG 720  
AATTTATGA CCAATTAAT TTTCAGAAAT GAATGAAA TTGCTTTT ATAGTAGGA 780  
GCGAAACAA AAAAAAGCT CTCAAAACA AAAAAAGCTC TGTAGATTC CAGCGGCTG 840  
ACCAATGAC TATGTCAAA GAGTGGCT GTAAAGATG CAGCGGCTG AAGCAGCAC 900  
TACAGTCTG GGGAGGCGG TTTTAAATC AGTGCAGC TGTGTCTGT GCGCTGTGAG 960  
TGTAGTCTT CACTCTTAT GCTTAGTTG AACTAAGAG TTGTAAACT TTCACTCTT 1020  
TTTTTTAAA TTCAAGAC TTGGAAGGA GAGCAATTA ATTTTGTAT TGAATGGC 1080  
TTGATG 1086

(2) INFORMATION FOR SEQ ID NO: 268:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1003 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double



5 GTTCTGTTG CTTCGACTTT AACAAATATTT ATACATATTC TGGGATATTA CGTATTTTGG 360  
CATCGATGCA AAGGACACCC TTTCGAACTT CCTGACCCAG GTTAAAGCAG GCTGCTTACT 420  
CATTTGGAAC AACTGGACTA TCGATGACAG TTTCACATCTT CACGGAAATTT TTTCACATTT 480  
TCTCCATATA TTCTATTTTT TCTGGAGT TTCTATACGA AGTATGATCC AACTGACTTC 540  
10 ATCTTAACA CAGCTTCTCT CCTGATGTTA CTAAATCCCA AATGTCOCA ACTACATGTT 600  
GTTCGATCT TTGGATTTTA TTAGTATGCA AATGTTTGA AACTGAAAA AATTTTACA 660  
GCTACTGAT TTCTATTAAG GAGGATGCG TTAGTAACT GCACTGTTTC TGTGATATG 720  
TGAATGACA AGTATTTACA TTGGAGGCC AATGCTGCTT CCTTCAAGTG CTGTTTGA 780  
GTTCAGATTT CCATTAATG ATGCTCTCTT TTATACACC TGTACATTT CTGAAGAGG 840  
20 GCTTTTAAG CAGCTGGCC AGGCCAGCT TATTAAGTTAA AGGCATCAC AGTGAAGGTG 900  
TAGTAGATA ATTCAGGAA ATAGAGATTT TGTAAAGAAC TAGGACCCAG TTACTTATA 960  
ATGAATGGCC ATTCGTTTA GAAAGACCA TTTCCTATCA TTCACTGTTG GTTATTTAAA 1020  
25 GCGACTTAC ATGTAAAGG GATCTCTC TATACAGTTT TATTAAGAT TATTTTATTT 1080  
ACCTACATA TTTCCTCTGT TTTATTAAG YGGATGATA TCTCTTGT TTATACAGC 1140  
30 CAGTTCCAC TATTAAGGAT ACTTTTGGG TTTTCTGGG GTTAATATG TGTATGTC 1200  
AATGAGCCA TTTTACAT TATTAAGTT ACAG 1234

(2) INFORMATION FOR SEQ ID NO: 270:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 574 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 270:

60 NGAAGTCCCT TCTGAGCCCT CTGTCTCG CCAGATGCTT TCAAGATATT ATTAATAACA 60  
TATGATCCC CATGAGCCC TATGACCA AAGTTTACA GGAGTTTGG ATGAGATGG 120  
GGCTGATGG GTTCATGCTT TATTAATCC GGGCTGCTGA TAAAGAGAT AAGCTTTGA 180  
AAGCTTCAG GCTGCTCT GTGCTACT AACGATTT ACTGGAGTA CATGTGAAG 240  
55 AAGAGTCAG TGTGCTGTTA AATTTACCA AGCCGTGTTA GATGGGAGC GTGAGAGTC 300  
ACTGTACACT TGTATAGTA CCGTTACTT CATGGGATG ATTAATGAT CTGTGATG 360  
CACTGTACC TGTACTGCT TTGATGTT TCCGCTCAG CCGTCCGGCG TGTACGCAAT 420

ACTCTGAGTA GATATTTTGT CATCGAGCC ATGCATACAG AATCTCACTG AGCCACCCAT 480  
CATTTGAAA TATTTACTTC AGTTGTACAG GACTTGCTCA TCGAGATCCA GGCATCTACT 540  
5 TGTATTTTAC TGTCAATATA ACUTTTATTA AACT 574

(2) INFORMATION FOR SEQ ID NO: 271:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1731 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 271:

20 GCTCCAGGTT GCGCTCTGTG CCGCTGACGA TCGAGCTCAC TACCTTGGA AATCTTACAC 60  
CTTCAGCAC TGTGTTTTTC TGTCTGTATA TCGAGGAAG GTTCAGACCA GGCATCAGT 120  
25 ATTTTGGGA TATTTATGAC GTGGACACGA GATTTTGA AGGGGCCCCG ATTTTGAAGA 180  
TTCTGTTTAT TGTACAGAA CAATACCTTA AAGTCTCTCG GAGCAGGTTT CAGAAATTC 240  
ATTTTACGG TGTAAACTG GTACTTCCA AGACCAAGTT TTCAATGTTA TTACAGAG 300  
TAGAGCGCC ATTAGCAGAG ATTCCCGAG TCAAGAGTGT TGTATTTT GAGATGAAA 360  
CTCATGTG CHTCCACAA ACTGCCCTCG AGCTATTTG CCGAGGAGTC GAGTTTACA 420  
TTGTTCTGA TCGCACCTCA TCAAGACCA TGAATGACAG GATGTTTCC CTCGAGCTC 480  
35 TCGCTCAGC CAGGATCAT ATGACACG AGTGAAGCT GTTCTCTTC AGCTGTAGC 540  
TCAATAGGAC CATCCAAAT TCAAGAAAT TCAAGATCTA ATTAAGGCA GTGCTCAGA 600  
GTGCGCTCG CTTCGAAAG TATGAGCAT TTGAAGACT GGTATGCTAC TCACTGTGCA 660  
AGGACAGTCA GGTGAAGGAC TGTAAAGCCA CACAGCTCT TCTTATCTCT ACTAGATTA 720  
45 AATGTTAAG TCAAAAAGG CTCCTTTTTT GGGCTTCTTA GTGACTTAA CCACTAGAC 780  
CAITTGATA CCAGATTTA GTTAAAGCG TCNAGGCTT CCGTCTGTC TTACTTCTT 840  
TTTTTGTAA TGTGCTTTTA TTATTAATA AATATTAACA TGAAGATCC TGTTTTGTCT 900  
50 CTACTGTGA CTCTGATGCT ATCTTTCCA AGTGCAGCT CTGTGAGCT TTCTTTAAT 960  
TGTTCATTT AAGAAATG AGTACCAAC AATGATTTG GTTTATAT ACTGTAGAT 1020  
GTATATGTT TATGTGAT GTAGTCTTT TACTTTACAG ATTCATTTGA ATAGATATAT 1080  
TCAATGAA TTATACCCA GAGCTCTGAA TCTTTTACC CACTCCCTC ACAATTTGT 1140  
CCACTATG AGTTGATG ATCTATCCCT ACCAATGAT GTTCATATAT TACTATCTT 1200  
60 TCTGACTAT ACTGATTTCT TATTTTGTCT ACTATTAATA AATCTCTGTT AATATTTCT 1260

CTTTAACTG AAGAGATG GATAGAGG GTTTCAGT CCATATAT GTTCAGAGGC 1320  
TGTATTACA TCTTGAAGT ATGGCTGCT GATATCTTT ACCACATCT TGAATATTA 1380  
TTCTAGTGT CACAAGATT AGCAAAAGA TAAAGCTTG GTGAAATATC ATTATTAAT 1440  
GTTCACTGC TCTCTATAT TTTCTTCAC TACTCTCCA ATATGTAT GCMAAAATC 1500  
TCAGTAAGA TTCTGATTA TTAATTTGT GTCATTTGT TCTCTTCAT AAATTAATT 1560  
TCATAAATA CTTATAGAG GTTTTGAAA TTTTTCAAA ATATGTGAA TGTGAACTG 1620  
CTGTCTTTA TATTAAATA ATTAAAGAA ATGATATGT ATTGAATTA TTTTGCTC 1680  
CACAGATGC CTCATGAGT ATCTCTCCAG GATCTCAT ATTATTTAA G 1731

(2) INFORMATION FOR SEQ ID NO: 272:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1320 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 272:

CTCTTAGA AGAAGATG AGATTTCGG GGGCAGAG CATCTTCC GCTGCCCG 60  
TCAATATTA GTGAGGGG AGACACCTC CCCAAATTC TGAAGATTC TTACTTCCA 120  
CTAGGCGAGT AGCCGAGAG TCTTAGTGC CGCTTCAG TCACTGCCG CTGACGGAG 180  
CTGCCCTGC CATTTTGG TCGTTGTGG CGGGAGGCT GTTCGAAACA GCTGCAGGC 240  
AAGTGGCAGA GATTAATTT GTTTTGAAT TACTGATTA TGAAGATATC AACCATGTTG 300  
TGGTTTTTAT GCTGGGACA ATCCATTTT CTGAGGGAT GGGAGATCT GTCTACTTTT 360  
CTTATCTGA TTCAATGGA ATCCAGTAT GGAATCTCT AGGATTTGTC ACGAATGGA 420  
AGCCAGTGC CATCTTCAAA ATTTCAAGTC TTAATCTGG AGAAGGAGC CACATCTTT 480  
TTGAGGCAAT GAATATATC GAACTCAT CTGTTCCTCA GATTCGAAT TCACTGCAAT 540  
TATTAGAGAG TATGCTCTG CAGATCTCG TAGTATATC TCTGTATCC TCAATTAAT 600  
CAATCACTCA GTTCACAGA AAGATGTGG ACAATTTCTA CAATTTTCT TCACTATTTG 660  
CTGTCTCTCA GGGCAGATG AGACAGGCC CATCTGAAT GTTCATTCG GGAATATGG 720  
TTCTGCCAAT GTTATAGGC ATTTCTGTC TCAATATTA AGGCTTTTAA TAACTGATTA 780  
TCAATTTGT CTATGAATAT ATTCCTTTT TGAATTTAA ACAATATCT TTATTTGAA 840  
CATCAAGCT GATCCCAAT AAGATATGA CTATAGAGAT CTGATGAGA ACGATTTCTTA 900

CCCTAAATAT TTTTATAT TGTGCCAT ATGAATTT AAGACAGA AAATATATT 960  
GCTTATGTTT TAGGACAC TATTAAAGT TATTAAATTT TGTATATTT CATTTAGAG 1020  
TACCATAT TATAGTATTT TATACTGAC TGTATACATG AATGCTAAT GAATCTATTT 1080  
TCCACTTTC CCGTATTTA TAAATATTT TTTCACTTT GATATCTTA GAGATGAG 1140  
GATCTTAGG AAGATTTGT TGAAGATGT TACCATGTTG TAGCATGGA GAGCATTTGG 1200  
AATGACTAG GTTATGATTT GGCATATGTC TACTATATG ACCTGAGCA AATTTCTCTC 1260  
ATCTGCTCAT CTGAAATG AGAATATG ATGAAATTTG AATTTCTCTA GTCCTCTCTA 1320

(2) INFORMATION FOR SEQ ID NO: 273:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 515 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 273:

CCCTGAGAG GGGCTGCTT GCGAGCTTG GAGGATCTG GATGGGCTT GCGCTGATG 60  
GGCTGATGT GAGTACTTT GCGACATCT GCTGGGCTCA ACTTATTTT AGCCCTTCC 120  
TTGTGTTCT TATGAGAGC AGAGGAGGG TGGCAGATC ATGATATCA GCAATGATTA 180  
TTCCAGCAC AGCGCTCTG GAGAGGCAAT GAGCATTTT TTTGAGAAA TGTCTATAT 240  
TGAGCCAGA GGCATCATT AATTAATTC TCACTTTGT CCGAGCTCTG TTTATATGC 300  
CCTTGGCAG ACTCAGAGG GGCATGAGC GCTAGTATT AGTATACAG GAGCTCTABA 360  
GATATATGG TCCAGAGA CCGCGGGGG GTGAAAGCC TGTGAACTTA TGTCTACGG 420  
AGCCGAGGT CACACTTTGA CTTTCTTACC ATGGCTCTG TCTATGAGG TATATATCT 480  
GGTAAATAT TACAGAGCA GTCATGTCT ATTTT 515

(2) INFORMATION FOR SEQ ID NO: 274:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2995 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 274:

TGACCACTT AAGGATTTCA TGAAGAAAT AGAAGAAAG CCAATGAGG TTAATCTAGC 60

AGTACCAGT GGAAGATCA TCTTGCTGT CTCCACCC GATTGCATCA CACCATCA 120  
ACATGATC ACCATATC GAGTCTGCTT CAGAGATTC CTGACATGG CTAGACGA 180  
5 CAGAGGCT CTGAAAGG CTTTTCAG ACTGTGCTT ATCTGAGC TCTTGAGGA 240  
ACTTCTGCA TGGATCCTT GGGCTGAGC CAGCTCATT CAGCGGATC AGGAGCAT 300  
10 CCGCAGAC ATTGACCGT TTAAAGCTT TATGCTGAG CATTGACAT TTATGAGGA 360  
GATGACTG AAGAGCTG ACTGAGCG GTTCACGAG ACATACAAA GGAAGACAT 420  
480  
AGAGCTACT CAGCGCTT TCAGAGAA ATCCCGAG CAGGCGAG AATCCCTAG 540  
15 TGAGCAGC CTCTCTCCA TGGCATCTT TTCACATCT GAGCAAAA ACCAGGAT 560  
600  
CAGCAGCTT TCTGCGCTT GCGAGCAT GTGCTGTTA GCACTGAGC GCGAAGAA 660  
20 ACTGAATGAT GCTTGGATC GCTGAGGA GTTCAGAA TTGCGACT TTGACTTGA 720  
780  
TGTCTGAGG AAAAAATA TCGTTGAT GATTCAAA AATCTGAG TGAATGAT 840  
25 CTTCGCGC ATTGATAG ACCAGATG GAGATACA CTTGAGAT TTATGATG 900  
960  
30 CATTATGCA TCGATGTC CAGCAGAA GTTAGAGAT ACTGCTGAG CTGACATTT 1020  
CGACGAGT GCGATGCTT ACATGATTA TTATGATTT GTGCTGTC TTGATGCAA 1080  
CAGAGTGG TATGACAA CAGCGATC AGATAAATC GAGATGAG TTACAGACA 1140  
AGTGGTGA TCGAATGTC CAAAAGCTT TCACTGAG CAGATGAG AGAATAATA 1200  
35 CCGTCTTC CTGCGATC AGTTTGGGA TTCTGAGC TTGCGGCTG TCGATATCT 1260  
GCGAAGCT GATGTTGTC GTTGTGAG GATGATGTC CTTGATGAA TTTTATGTA 1320  
AATATATTC CTGCGAGCA CAGGTAGAA CTAAATTTA ACTTAGAG AATTCATTC 1380  
TACAGAGG AGCATCCAG GGAATGACC CTTTCCGCTC AGGGGTGCA AGTCCAAAC 1440  
40 CACTTCCG CCGAGCTTC CTTACTGTT CAGCTGAG TCGTATGAG AGTAAACAA 1500  
GCTATATC CATGCCAT TCTCAGCA CCGCAGCAG TCGAACAG GTTATCCAT 1560  
45 CACTAGTAG CATTGAAA CAGCAGAC CACTTTTCA TTCTAGTGG ACATCCCTG 1620  
CTGATATC CAGCATTA TCTTCCCG CCGTCCAGG GTGCCAAAC TATTCGACA 1680  
50 GACCTAAA AATCTCCAG TGGCTGAG AGTCGGCTG GAGTCTGAG CCGGAGTCA 1740  
GCGAGAGC GCGAGGAG TGAAGCTCT GACTTTGAC TCTTAGAG CATTGCTG 1800  
TTCCAGCT TCGAAGCA GGGTGAAG GGGCAGGC AACTTCAGA CAGGGCTAAA 1860  
55 CAAAGCTTC AATATCCA CATTGTAA GAGGACCC ACTGCTCC CAGGACTCC 1920  
AGTTCAGG CATTACAT GTTAGAGC CCGCAGCA CTATCACT TGAATCTCC 2000  
TCCATACAT GGTATAT TATTTGAA CCGGAGAT TATATGTA AATGTAAA 2060  
60

AGATATTC TGTATGAG CTGCTTAT TTTTCTTT TTTATAGTA CTATTTCT 1920  
GTGATATTT ATGTAGTAA AATTTCCTC CTGTAACCC TGTATGAT GGGGCCAG 1980  
5 AATGAAAT TTGAGAAA CAGTGAAA GGTCAAGTA CAAATGTGA TTAAMAAA 2040  
AAGAGCTAT TATAGGCTT TCTGCGCTT GCAAGGCTT AAGCTGCT TATCTTTAG 2100  
10 GATTATCTT AATGCACT TCTTATAAA CTTGACTTC TATCTGCA AGATATTA 2160  
TATTAATA ATAGAAATC TCACTGCTT AAGAACTCT TTTTGTAA ATCAGACA 2220  
CTCATTTAG CAGAGTCA GGGAGGCT TTTTCATG TTTATGTT TGTATTTT 2280  
15 AGCTAAGAG AGGAACTC ATCTAGTAA CATTGACA TGGATAGC AAGAGATTT 2340  
CATTCATA CTGCTTTG ATATGTTTC AGTACTGCT GTTTAAGCA CAAATAGCT 2400  
CTAGATTA GGGTAAATG TAGTGTCA GAAAGCTCA GACATTTG GATTATAC 2460  
20 TGAATGTC CTGCTATCC AGCTAGCA CCAATAGTC TTGTGTTAC CAGAGCCAG 2520  
ACCTTGCA AGGATAGC TCGTTGCTA CATTGTGAT TTGATGTT TTTATGAC 2580  
25 TTTTCTCT ATTATTTA ATGTCATC AACTTCCAC AACTGAGC ATGATTTCA 2640  
CGAGCTGTT CTGAAATGT GCACTAGCA CAAAGCTG CTGCTCTTT AATGAGTTC 2700  
ACCAAGAC TTGTAACTA GTCTGTTG CTTTCTGCT TTTTGTGG TATATAGTC 2760  
30 AACTGACA GTGACATA AAGGGGCTG TCTGGGCTC CTGTTTTTA CCGTCTGTC 2820  
TTCACTCG ACCATGTC TGTGTGTTA TCTCATGTC TTTATTTCA GCGCAACT 2880  
35 GAGCTCTAC CATGAGTC TTAGAAACA AGACACTT TTGTATTA ATTGCTTCA 2940  
GTAACTAAA AAAAAAAA AAAAACTG AGGGGGCC GGTAC 2995

## (2) INFORMATION FOR SEQ ID NO: 275:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1990 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 275:

GGAGCCCGG CGSCTCCCG GATGCTGAG CAGCGGCTG CTGCMGCTG TCTGCGTCA 60  
ACCGCAGAG ATGAGCTGC TCTGGGAT GCGCTTGTG GCTTACGTG CTTCTGTTG 120  
GGCACTTC GTTATATCA GGTCTATCA GGAATATGT GACTAATA TTGAAAGCA 180  
GATTGAGAG ATGTTGAC CACTAGAG GAAATACGA GATTAGAA AAGCTTAC 240  
60 CCGAATAC CCACTATA AGTTTATC AGAAGAT CCGAAGAA TTTTGAATA 300

5 CAGAGGCGC AGGTTCTGTC TAACTGACA AACTCATGAT GAGCGCCAC 360  
GAGGTGACCG TGGTGGACAA TTCTCTACG GCGAGGAGA GAATCGTGGG GACTGTGATC 420  
GGCATGAGA ACTTCGAGTT GATTAACGAC GACGTGTGGG AGCCCTCTTA CHTGAGGTT 480  
GACGAGTAT ACCATCTGCG ATCTCGAGCC TCCCTCCAA ACTACATGTA TAACTCTATC 540  
600 AAGCATTTAA AGCCGATPAC GATTGGGACA TTAAACATGT TGGGCTGCG AAAACGAGTC  
660 GGTGCGCGTC TGTCTCTGCG CTCGACATCG GAGGTGTATG GAGATCTCTA AGTCCACCGT  
720 CAAATGAGG ATTACTGCGG CAGCTGAT CCAATPAGAC CTGGGCGCTG CTACGATGAA  
15 780 GGCATAAGTC TTGCGAGGAC CATGTGCTAT GCTTACATCA AGCAGGAGCG GTTGGAATG  
840 GAGGTGCCA GAATCTTCAA CACCTTTGGG CAGCGCTGCG ACATGAGCA TGGGCGAGTA  
900 GTGCGAACT TCATCTGCGA GCGCTGCGAG GCGGAGCCAC TCAGGTATTA GCGATCGCGG  
20 960 TCTGAGACA GGGGTTTCCA GTACTGAGC GATCTAGTGA ATGGCTCTGT GCGTCTCTGT  
1020 AAGCGAAGC TCAGAGGCGC GGTCAAGCTG GCGAACCGG AAGAACACAC AATCTGAGAA  
1080 TTTCCTCTGT TAAITAAAAA CTTTGTGTGT AGCGAAGTG AATTCAGATT TCCTCTCGAA  
1140 GCGCAGGTC ACCGACGAAA AAGAAACCA GACATCAAAA AAGCAAGCT GATGCTGGGG  
30 1200 TGGGAGCGCG TGGTCCGCTT GCGGAGGCT TTAAAGGAG CAATTCAGTA CTTCGTAA  
1260 GAATCTGAGT ACGAGGCGAA TAACTGATAC ATCCGCAAC CAAAGCTGCG CAGATTAAG  
1320 AAGAGCGCA CTGGCGACAG CTGACTCTCT CACTTTTGG AGCAGAGCT ACCATGTAC  
35 1380 ACTTGATGGG ATGTATTTTT GCGTTTTTTT TGTGTGCTTT TAAAGAAAG CTTTAAAGCG  
1440 TGTCAATGAG AAGAACTCG AATTCATGTC TGAAGCTTGC TTAAATGAAA TGGATGTGCG  
1500 TAAAGGCTCC CTCGAAAAA CTGCAATTTT TGCCTGAC TTCTTGAATC TCCTTTTTA  
40 1560 TGTAAATAG CGTAGATGCA TCCTCTGCTA TTTCAGGTT TTTTAACTT GCTGTGAGG  
1620 CATATGTTGT GACTGTGCTT GACAGTTTTA TTACTGCTTT TCCTTTGTGA GCTGAAAGG  
45 1680 AICNTTAAGC GCGACNAAA ATGCCGATTT TATTATAAA AGTGGTACT TAAATATGA  
1740 GTCTTTATAC TATGCTAAA GAAAAATCTT AGCAGTATG TCAGGTGCTG GTGCGCGCGC  
50 1800 ATTGATTTTA GCGCAGTAA AGAATCTG TGTGAGGCT TTATGTTCT CTTTAAATC  
1860 AGAGTTTTTC CAGGTCTTAC TTGTGAGTGG CAAACTTGAC TTGTGAATAT TCCTGTGCTT  
1920 CATGATCAG GATTTATGAA ATCACTACTG TGTTTTCTG GATATCTGCG GCGGCGCGAG  
55 1980 GTTGGGCGCG ACAAAGTTAA CATATCTCTG GTTAAACATG GTTAAATATG CTATTTTAT  
2000 AAAATATTGA

## (2) INFORMATION FOR SEQ ID NO: 276:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2436 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (11) SEQUENCE DESCRIPTION: SEQ ID NO: 276:

5 AACCTGCGTT AGCTCTCCAG GGTAAAGCG GTGAGCTTT AAAACAGAA GAGAAAGAA 60  
15 TTTAAAGTCC GTTGCATGGA AATTAAGAAA CATATCAT GTTTTATCA AGCATCTCT 120  
CGCATTCCT CTTCTTTATA AGAGCACT AGACTATCC TGGAAACCTG TACAAAGGT 180  
20 TCGATGCG CAAAGAGAG CAGTGAGAA TACAACTTCA GTTTCACAC CCAAGAAATC 240  
TTACAGACA CAAAAGAG ATGCGAGCA GATTATTAAC CTTCCCACTG GCAATATPAG 300  
CAGCAATTC GCGACTTTTA ATTATGACA GAGAGAGCC TTCAGGGGAA GTAGAGGTGG 360  
25 CCGAGGTGG GGCACAGAG GAATCTGTA TCGGGGAGA CTCTACTGAA TAAAGATCA 420  
GCACTTCTCA GCACTGTGAT GAGCTTAATA TACTTAATTT CTACTACTCA TTGATTTGCC 480  
GGGATATCC CTTTAAAGAG ACTGCTGCTT TCACTTAAA ACTTAATGTT CTTTATCTT 540  
30 TTGTATGAT GACCTACTTT TGTAAAGAG CATGTTGTT TCGAGGTAA AACCAAGTGT 600  
ATATTTTGG ATGCTTTGTC TCGAATCTTG ACTGTGTTT GCAATATCAT TATTCAGACT 660  
35 TCAATTTGG AATCTTTTAA ACATCTGAT AATTTGTTGT TGAAGCTGT TCATCTTAAA 720  
ATGTAAAGAA ATTCAGTCTA GTTCTGCTGA TAAAGATCAT CAGTTTTGAA AGCTTACTGA 780  
TTTTGCTGTT CCGCTGTGAT TTTTAAACA ATATATGAG AGAGTATATG GTCAATCTTA 840  
40 ACATTTGTT TTAATTTGTT AATAAGCTG CTGGCGACTG GTGCGAGTCT CCACTACTGT 900  
GTCAATAAG CAAATTACTT ACATGCTTT CTTAAATAT AGGATATGCA TTACTATTTT 960  
45 AGGAGAAAT AGTGTGCTTT GCAACCGCTA CTTAATCTTT TTCCATATAT TGTGATGAA 1020  
ACTTTTGAT ATGGAATCTT ACTATTGAA TAAAGATGTT TATGTATAT ATATGATAT 1080  
50 ACATAAGAT ATATGTGTGT GTGTGTGTGT ATATATATAT ATATGATAT TGTGATAT 1140  
GACTACACA CATTAATCAC TTTTAAATTT CAGAGAGCG GTAGCTGAC AGGTGATTA 1200  
TCTTTTGGG GCTGAATCGG TTAATTAATT GTTATTTAGG TTTTAACTCC CAGTACGAG 1260  
55 GATTTCTTAG TTAGTTCGAC TTACTGAT ATGCTTATTT AAATCTTACA ATAAAGGCTG 1320  
CATTTTCAA GATTAATGCG AATGCTGTT GGTCAATTA CACCAAAAT ACTGATCTG 1380  
ATGTACTATC AGTTTCTTAC AGGAGAGAT GATTAATTTT ACATTTGGA GATTATGAA 1440

CCAGGCTAC CCGAAGAG TGACTGTATA AGTGTGACC ATATAGTAG GATCTCTC 1500  
TGGTTTCTCT TTTCGACTT TCGAGTTT ACTTCTCAG GTTATTATTC AAAAAATTG 1560  
5 TATATAGTAG CCAATAGAT TTATTAGTTA AACACACAGA TGGGGGTTT GTCGATGTT 1620  
TAATGCTG GCAATTTTAA GTACAGAGA CCGTTTGTTC TGCATTTGAA TGTTTGCTAT 1680  
ATTTTTGT TTCACTTAT CTTCGCTCC CAGTTTCTCT ATTCGAATCA ACTGCTGAA 1740  
10 TCACTTCT ACTAGTCTGA TGTATTATTC TCGAGATAG TTGTGATTC CAATGAGGT 1800  
GTCTGATTA CCAATCTCT TACTGCTGAG AAGAGCAGA ACTGCTTAT TGAATTAAT 1860  
15 GCACTGTGT ATGGGGAAA TAGTCTGAA AGCTTAGAT CATACAGTC AGCAAAAGTT 1920  
GTCTAGCTTG GCTATGAGT GTTGCAATA ATCTCTAAC ATTCCAAAG ACCATGAGT 1980  
20 GAACTTAAC TCCCTTGAA TCTGAAGAA GGAATATATA ATTGCCATTT GAAACTGAC 2040  
CAGCTAATCT GCACTCTGA GATGATCAG CAGTGGCC AAAGCAATTT CAGTACAGA 2100  
AATATAGAG ACTACAGCTA ATTAATTTTG AACTATTAAT ATAAATTTAC CACTTTTGT 2160  
25 CTTATAGCC ATATTGTAAA ACTGAGACT GAGCAGAGT GACTTTACTT TCTCAGTTT 2220  
GATAGTATG TCACTGTTC CTATGCTTC ACCCTTCCC TTCCCTTTTC TAGGCAATTA 2280  
GTGCAACT TAGGTATTT TTGCTTGA ATTTGATGA AATCAATTAAT GCAATGAT 2340  
30 TTTTCTTTT GCAAGACCC TGTATATAT CTGTTTTAAA TGTAAATGTC CCGTTATCT 2400  
TTTGAAATTA ATTGCTTTT GTAAATTTAA AAAAA 2436

(2) INFORMATION FOR SEQ ID NO: 277:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 782 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 277:  
GCACTAGCT TTCTCAGCC TTCTGTCTCC CCAATATAG TTTATTTGAT TGTCTGAC 60  
TCACTCTG CTTTATTA ATTTCTTAG GCGCTGAG AAGCAATTC TACTGAGAG 120  
GATTAGAGC ACTTGAGCA GCGCCGACCA TCCAACTCT GCGATTTGTS GTGGAGGAG 180  
GCACTCTG GGAATAGAC CAGCAAGAT AAGAGAGCT CAGTGTGAA GCGAAGCTG 240  
TGCAGATTT AGTAGTCCA AATGAGTTA TATCCCTTCC CCGTTTAT CAGATCTG 300  
TGAATGGA AAGACAGA AGAGAGGAT CAAAGATAG TGTATGACA TGTCTCCAG 360  
GCAAGGAGT GGTGGAGTC AATCCGCGT GACAGCTGG TCGAGAGCCC TGTTTAGGT 420

TGTGCTGAT CCGTCTCTG TATTAGTTTT TCCCTGGA CAGAGAGCC CTAGAGAG 480  
GGGACTGAG GGTCCGCGG GATCTTTTC TCCCTCCCT CATTAGGCA GAGCAGCT 540  
GCTGCAAC CCGTCCCTC AAGGAATGCG TTGCGGAGC AATGCCACC AGCATACCC 600  
TCTCTTTT TTTAGTCA ACTCTGTTT ATTCTTGGC TTGCTGCTT CTTTCTCC 660  
10 CTCTAGCT TTACTTCTGA TTTCTATTC ATGGAATTT GATTTGAGT TAACTACA 720  
CAATGCCCC AACACAGCT CTTCAGGAA AAAAAATTA AAAAAAAA 780  
AA 782

(2) INFORMATION FOR SEQ ID NO: 278:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 961 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 278:  
GAGTCCGCG TCGAGAGCG TGTCTGGGC GGGGCTTC ACCATGGCT CCGCAGAGT 60  
GCACTAGCC ATCGAGATC CCGATAGCC CTCTGGAGC CAGAGAGCA GCGCAGGCC 120  
AGTGTGAG GAGGAGAAA CTGCGAGCC TGTGTGAT TTCTTGGCT GGGTGGCTG 180  
35 CAGAGCAG AACCTTCCA AGTACATGC CATCTACAC AAGAGGCT GCATCTAAT 240  
CGATACAGA GCGCGCTGC AATGCTTT CTCTCGAG TCACTGGTA TCCCTTACT 300  
TGTGTTTT GCGCAGAGC TGTCTAGCT GTCTTTTGT TATGAGTTG AGAGAGGCC 360  
CCTCTCTTC CATGTCTTA GAGAGGTG GTTATCTG TACGCTTAG TGTGAGCT 420  
CCTCAGACC CTTGCTTCT GCGCGCTGC TGTGTGGC ACCATCTTG ACAGGCTCC 480  
45 TGTGTAGAG AACGTGTAG GGGTCTGCG GCGCTGACA GCAATCTCG AGCGCGGCG 540  
CGCATGCTG GCGCTTTC TGTGTGGC CTTTGGCTG GTGTCTGCT TGTTCAGCT 600  
CCTCTTCT CCAATCAG CCGTCTTCCA CACCACTTC TATGACAGC TACAGAGC 660  
50 GGGCTCTGC TGGCGGAGC TCTACTCTA TCGAGGCTT GACGAGTAG TCTTGGCAG 720  
AAGATAGAA GCAATGTG AGGCAAGCT GCAAGCGCG GTCTGTGCG GTTCTGTGA 780  
55 TTTCTGTCA TCTCAGAG TCGAGAGCT CCGTACTAC CTTACTTACT ACAGAGCT 840  
CTGTGTGAG TTGATGCGA ACTGCTGCG CTGCTAGGC CATCTCTCA TCTCACTCT 900  
GCTCAGAA TAAATGCTG AATCTGCC ACNAAAAA AAAAAAAA ACTCAGAGCG 960

506

507

6

961

5

(2) INFORMATION FOR SEQ ID NO: 279:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1228 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 279:

15 CGCGCTTTTC AGTTCGGTCT CTGTGCTTAC GGCCAGGCC AAGTAGGGGA TTGCGTTCC 60  
TCGATGCGA GCGCTATCAG ATTGGATAT GTCTTCTATA TTTCATTGGA TTTCACATGG 120  
20 TTTCACAGAT GTCTACAGT TTTTAGGATT ATATAGAAA ACTGCTAAC TGGTATTCT 180  
TGGATTGGAT AATCCAGAAA AACCAACATT GCTACATG CTAAAGATG ACAGACTTGG 240  
ACACATGTC CGACATATAC ATCCCACTTC CGAGGACTG ACCTTCTGCG CATGAGCTTT 300  
25 TACACTTTT CATCTGGGTS GCAATGTTCA AGCTCGAGC CTGTGGAAA ACTACCTTTC 360  
TCTATCATAT GCGATTGTAT TTCTGTGGA TTGTGCGAG CAGGAGGCG TTTTAGAGTC 420  
30 AAAAGAGAA CTGTATTCAC TAATGACGA TGAACCAATT GCTAATGTGC CTATACTGAT 480  
TCTTGGGAT AAGATGACA GACTGAGCC CATCACTGAA GAGAGCTTGC GAGAGATTT 540  
35 TGGTTTATAT GGTGAGCAA CAGGAAGGG GATATATCT CTGAAGAAC TGAATGCGCG 600  
ACCTTAGAA GTTTTCATGT GTAGTGTGCT CAAGAACA GATTACGGAG AAGGCTTCCG 660  
CTGATGCGA GATATCATG ATTAACAAA ACTACATG GTTCCAGGTC TCACAGTTTCA 720  
40 GCTTACTCA GAGATTGTAT TGTTCACAT GCNTACTTG AATTCAATAG ACTTTTCTG 780  
GTATTAAGAC AAGATGTTTT TAGATTATTA ATATTAAATC AACTTAATTT GATATGAGAT 840  
45 TGAAGAACTGA TTCAATAG TTTGAGTATC ACATGTTAG CTTTCTAATT CCATTAAGAT 900  
ACTTGTGTTT PACATGTTTAT AATGTGACAT CAGCCGAGG GCKTTTGTAA AGAGCACTT 960  
TCCAGCGTA CATTGAGGC ACTTTTAC ACATGAAAC TATTAACCAT ATTTAAGGC 1020  
50 TATCATGTTT AATTTTTTTA TGTACTTTTC TGGACTAGT TTTTAATTTT TGAATATAT 1080  
GTCCACCTAT CTTAGTCTTA CATTATATTA TTAGCTTATT CATTGATTC ATGATGCTT 1140  
1200 ACGATTTC AATACCTTTT TTCTATGCA AAGTCAATGC AATTAAGAA ACTCTAATGT  
55 TTGCAAAA AAAAAAAA AATATCA 1228

60

(2) INFORMATION FOR SEQ ID NO: 281:

(2) INFORMATION FOR SEQ ID NO: 280:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1327 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 280:

10 TTCTGGGTCT CCGGACAGGT GAGCAGCTG ATGAGGCCA CGTCTCTTAT GCGGCAGCTG 60  
GCGGGTCCA GCGAGTCTG GCGGCTCTCC GCGAGGCGG GCGAGCGCG TTACAGGTGA 120  
15 TTCTGATTT TTACATGACC TTGAGCAGGT TTGCATATTA TCGAAGCCA TCGCTTCTTT 180  
CTTACATAT TCTGTATAT ACACATATCA TCAGTGAGGA GTTGTGAAA GAGCTCACAG 240  
GCTCTTCA CCACTATPAC CCATATGAGA TCGACCCACA CCGGACCTTC AAGGAGAAC 300  
20 TACTCATAT GTGTGAATGG TGGACCAAG CCGACATCT CTTATGTGAG CAGAGATTC 360  
AAGATTTC GATAGCCGAG GTGTGTAGAG ATTCATATGC AATGCTCAG GAGGATATTA 420  
25 AAGCTTCTT CACACACTC TACCATACA AGTTTCCCTT TTCTATCTTT TCTTCCGCGA 480  
TTGTGATAT CCTGAGAAA ATTAATCCAC AATGAAAGT GTTCCAGGCC AAGATCCACA 540  
30 TCTGTCTAA CTACATGAT TTATATGAG ATGCTTTCT CAGGGAATTT AAGGCGCAC 600  
TATPACACAC ATPACACAG AACAGCTCTG TGTGTGAAA CTGTGTATAC TTCCAGCAAC 660  
35 TTGAGGCCA AACCATGTC ATCTGCTGG GAGACTCTAT CCGGACCTC ACCATGCGCG 720  
ATGCGGTTC TGTGTGTGAG AACATTTCA AATTTGCTT CTTGAATGAC AAGGTGAGAG 780  
AAGGCGCGA NCGCTAATC GCACTCTTAT GACATGTCG TCGAGAGCA CAGACTCTG 840  
40 GATGTGTCA ACGGCTACT CGAGCAATC CTGTGCGAG GGGGTCCAG TCGAATGCA 900  
AAGGCGCTGA AAGGCGAGGC TCGAAGGCC SCTGAGGCC GTGTGAGCA GCGGCGCTC 960  
CCGAGATCT GCTTCCCTCT GAGCAGAG CAGAGCCAGS GTGCGAGCA GTGCTGCGT 1020  
45 CTTTCCGCG CCGTCCGTC TCTTTTCCCT GAGCAGCTTC ATCACAGAG GCTTGAAGCA 1080  
ACCGGCAAT GTGTGAGGCC AAGGCACTG TTCTGTGTA ACTTGTGACC ACGATATTC 1140  
50 AATGCTTAG GCAATGTCTA CTCCAGGAT TTCTTCAAA ATTTTAAAC ATGCGAGTT 1200  
CAAGAAATA TATGTGTGCA AACGATCA AATTTTAA ATGAAGAAA AGCTGCTCTG 1260  
1320 ATTGAGGGA TGTGTGCGS GTTAGAAGCT GAGCTCTG GCTTGGGGG ACATGAGTG  
55 CTTCTAG 1327

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 799 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 281:

10 TCACTCTGCC TACAGCTGAG AGCTGAGATG ACTGCGCCCT CAGCGCTAC TGTGACGAG 60  
TGTATATAC AGCTGAGT AGCTGAGG CAGCGCTGAG GGTGTGCCC GTCACTGTGT 120  
GGCTTTGGCT GAGCTGAT TCCAGCTG CTTGTATTTT GAGAGAGACT GAGAGAGCA 180  
TAATCTCTGC CTGCTGACC CAGCTGAGG CTGAGTCTC TGTGAATACA TTATCTTGG 240  
ATGTGGGTTT ATTCAGGCA LAGACATTTT AGTGGCTGT ACTGATTTG TACATATTA 300  
TAAATATCTA TTCAAGATT GTTCCATTA TGCAGTGTCT TTGCTCTGG TACAGCGAGA 360  
GGCTTCAAC CCGACTTGG ACTTGAGGAC CTAGCTGAG GAGCTTTTC AGTGTCTCT 420  
AGAGAGGAT TCTGATATCT AGCTGCTAC GAGATGTTT TCACCAAGGT CACAGAGCA 480  
TTGCTGTGCT GATGGGTTG AGTTTGGTT TGGTTCTGT TTGAGCCCA TATCTAGAGA 540  
ACATTTGAAA CAGTCTGAC CTTTGATAG GTATTCGAT TTCAAGGCA CCAATCCATT 600  
TTGTGATTTT TATGTCTG TGGCTTATA ATCAATGTA CAGCATATAT AGCTTTTCT 660  
CCATTTTCT TCGAGGAAC ATACCTTAG TTTTATTTT TTGTATTTG TTTTATTTT 720  
TTTGTATTTT CTTATATGAG AAAAATATA ATAGTCAAT TTTTATATCT AAAATATGGA 780  
CAAAAAAGT CAGGGGGG 799

## (2) INFORMATION FOR SEQ ID NO: 282:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2196 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 282:

50 AAGACTCTA ACATCATGA GCTTGACAT GAGCAAGAG CTACTGTGC CUSCGATG 60  
GCTGAGGCT TCGTACCT CCGAGTGA TGGTCTCT ACTACAGCA GCTGTGTTT 120  
CTGCTGGA TGGTCTTGC TTGCTTTAT ATGACTGTGC TGGCTTTGA CTGCTAC 180  
ACAGGTAGG CTACACTA GAGAGTGA GTTTCATTC TCAATATTT GATGGAGCA 240  
TCACTATA CTGATATAT GGAAGCTTA GCTTTTACT GCTTACCTG AAAATGTGT 300

TTGCTTGGGA CAGTCTGTAT CTGAGATTTG GCAAGACTTT CTGTTTGAT CTGTGTGTG 360  
ATCTCTGTAT TCAATGCTGG AAGCCCCCTG GACTTGTGG TTCTCTCTTT TCAAGATATC 420  
5 CACTCAAGGT TCAITCAAGG AGATGCAAT ACACCTACCA AGATACCTGA AATTCAACT 480  
GAATATACA TGTCTAATGG GTCTAATCTCT GCTAATATTTG TCCGAGAGAC AAGTCTGAA 540  
TCTGTGGCA TATCTCTGT CAGTCTGCTG TTTCAGGGG TCAITGCTGC TAGAATCGGT 600  
CTTTGTCTCT TTGATTTAAC TGTGACAGG TTGCTGCAAG AAAATGTAT TCAATCTGAA 660  
AGAGGCAATA TAAATGATGT ACAGAACTCC ATCAACTATC TTCTTGATCT TCTGCATTTG 720  
15 ATCAATGCTA TCTGTGCTCC AAATCTGTA GCTTTTGGCT TGTCTGTATT GATTTCAATC 780  
TCTTTTGGG CAGTGGGCA CATATGTAT TTCCGATTTG CCGAAATATC TCTGGGAAC 840  
AAGCTCTTGG CTTCGCTCC TGAATCAAAA GAGTTTGA GAAGAAATCA AGCAATATCA 900  
TCTGTGTATT GAGACAGTTT AACTGTGCT ATCTGTATC TAGATATAT AGAGCAATG 960  
TGTATTTT GTACTGAGA ATTCCATTA ATGCTGGGT GTTTTGTCT GTTTTATCA 1020  
25 CAGCTGCTC TTGAGACTA AAGCTGTATT AGAAACTTA ATCTAGGCA AATTATCTGA 1080  
TTAATTTCCC TTATTTGAG GATGGA AAAATGGA AAAATGGA AAGAAACT CAGTTTAT 1140  
AGGAGACTA TATATATAC ACTGAATCC GCTATTTCT ATGATAGAT AGAATCTTAC 1200  
30 GTAAAGAGT GGTATGTCAC GTGATTCAG TTATCATTTG ACAGTCTCT ATCTGTACTA 1260  
GATTCAGAT ATGCTAGTTT TCTGCAAAC TCACTCTGT TCAAGACTAG CTATTTTAT 1320  
35 TTTTTCATC TTATTTATTT TTAATACAA ATTCTTCAG TATGAGACT AAATTTGAT 1380  
AACTATATAT ATCTTATG ATCTATATG TCTTAACTA TTTACATGTA TGTGGAAAA 1440  
CAAAAGCTT AACTAGATT CTCTAATAG GTTATGCTT TAGCTTAAAG AGCACTTTG 1500  
40 TATTTTATAT ATCAGATGG GCAATATAT GTATGAGCA TATGTAGCAG TTCAGAGCT 1560  
GTTATCATG TAACTGCGAG GTAGAGCAA AGCTGTAAAG TAGATTTATC ACAGATGAC 1620  
45 TGCATACAGA CTTCAATAT GTCAATAGT TGTGATAGA ACTTGAAGC CAAAGGCCAC 1680  
ACAGAGGCC AGAATGCCA ATTTAACTA TGTATCATC ATTAGTATC TGTGTGTAG 1740  
AAGATGAGG TGTAGGCTT CAGCTGCA AGTTACATGT AGAAGGCCA CACTTGTGAA 1800  
50 GGTTTGTATT TACAAATCA TGTATTTAC ACCTCAGGT AGATATTTT TATTTTACT 1860  
GTTTTATAC CAGAGTTAT TTCTACATG TTCTACAGCA AGATATTTA TAAAGATATC 1920  
55 CTTTTCAT GCTTTGAGA AGATAGAG AAAAAGTT TGTATATAT TTAATATAT 1980  
GTTTTAAAG TCAATTTCA ACATGTCTT ACCAGATG TACTTTGCT TAACTTTTA 2040  
TATGCTATT CAGTGAAGT GCAATGCTT GCTATGAGCA CTTTCTTTT CTTTGAAGT 2100

510

511

TAACTCTTGG CTTCATCTTT CTACAGTATG ACATATGAT TTGCTATGTT GTAAATCTT 2160  
TGTAAATAAT TTCTATATTA AATATTTGAA ACTTAA 2196

5

## (2) INFORMATION FOR SEQ ID NO: 283:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1185 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

15

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 283:

GGAGTTAAGG CTTCGTATTA GGAAAGAGAG TCTGACAGA GCACACAT CTGAGCTCC 60  
AGGAGTGGGG GATGCAGCAT CAGATTCAT CTTCGATTTTC TGTAAATA CTTTGTACTC 120  
ATAATGGATC TCAACAAAGA TCTGTATTTT ATCTGTGGCT CCATCTTCCC TCTGGGTCAA 180  
GTAGATGTTA AGCTGGACT TGGCAGCCT CTTAACATGA AGAGATCTAG CTAGACAGAC 240  
AGACTCCCC ATTATGGGA ACAGATATC AATTATCT CTATTTATTA AACATTTTTT 300  
TAAAGTGGCT TGGGTATTA AATCTAATG TCTGGGTGT GATCATCTAG GAGCATTTA 360  
CTATCTACT CTGATCTT TGTCTACTG GAGATCTTT GGGGGCTGG AGTCTCTCT 420  
GTCCAGCT AAAGGAAG CTTCACAG GTTAGAGCCA CAGAACCTC GCACAGAG 480  
GGCGGTGAG GAGATGAT GTTACAGGA GGAAAGGAG GAAAGGGGT GCACAGAG 540  
TAGAAGGCAA GGAAAGGAT CCGACTGGA GACCGATGG GACACTTAA TTGTGANGA 600  
GGAGGATCT TCTTCTTGA ATCTGACA CAGCTAGCT GACCTTCT TCGAAGTCC 660  
AGCTGTTTC CATTGCTAG GGCACCTCT CCTGCAAG CAGCAATGT GCTTCTATC 720  
AGGAGGAAA AGTATCTATC ATGTGACAA GAGTCACT TCGACTTGC ATGACTCTT 780  
TCCAGGCCA CAAGAATCC TGGTAGAGT GAGGATCCC TAGTCTTAG GCTGTCTTT 840  
TATAGAGTA CGATGACAC ACTGCTCTA GTCTCTGAA TACAAAGC ATTTGAGTC 900  
CATCTGTCA TATGAGCTG TTGAGTTTT TCGAGCTAA GTTCATGAC CAGCATGAA 960  
GGGAGTCT GAGGGCAAG TACAGCTG TCGAGTAT GTTCTTTTG ACTTATGCA 1020  
ATAGAGGTT GACCAAGCCA GCTTGAAGA CGAAGTGTT GAAGACATG AGGAATATTA 1080  
TATGATTT GAGAGGACT GGAATCTCT GATCTACTG AGGAGAGGY TTCTAGTTC 1140  
TTTGATGAG AGCTGTTTCT CTTCCTTAA GCGAAGCAC TCCA 1185

60

## (2) INFORMATION FOR SEQ ID NO: 284:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1634 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

5

10

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 284:

AGCGAAGGCG GAGGTAGCG GAGGGTAGC AGGTGATTC CTAGGGCTCG AGGTTTAGC 60  
AGGAGCTGCG TCGATGCCCC TGTATCAGAG ACAAGCCAC GGTCTACTCG GTTGGCTAGC 120  
AAGCTTGTTT TGTACAAAG CAGGTGGGA GTCTATTTTT GTACATGAGA TACATCAGAC 180  
TTACTGTGCG GCGATATTC TGAATGAGT CTGATGTTT TACATGATG CTTTCCCTGC 240  
CCACACAAA TTGTATCAT AGTCTTCAGA TGAATACAC CTTTTCGCA GCTGCCAAC 300  
AAGAGCTGCT TCTAGGCTCG TGTATATGT CATATTTAGC STTTTATAT ATGACTTTTG 360  
ATTCTGTTCG TTGTATTTTT AGCAGATGT ATGCAGCTTC ATTAAATAT ATCTGTGTGC 420  
ATACAGATAC GATATATAT GTTGTGATAT GCAATATCT CTATATCTTA GTTTCAGAA 480  
GTTCAGCTGA AGCAGATGGA GTCTGACGC CAGGAGACA CCGTCATCC CTGCTATTAG 540  
TGTTTGCGAC AAGTATATAT GATCTCTCT TATTAATAT TTCTATTTAG AAGACTGAG 600  
CAAGAGTAT AGTCTTCTCT GTTCTTTTTC CAGCTAAGTG AGGATCTTTC GATGACTTGC 660  
TGTGTCTCT AAGCTGACT TTGGGGGAT CTCTGAGTA TTAGGCCCC TTTTCTGTGC 720  
TGTACTCTCG TCTGTGCTCG TGTGTGTG TGATGATC TCTTGCATCG CTTGCATGTC 780  
TGTTTGTGCG CATTTGGGGA TAGGTGCTCG AGCCGAGAG ATTTGCAAT TTTTTGAGGC 840  
CTGCTGCCA ATGATAGATC ACTGCTGTG AGCTGTATG TCTGCTTGT TGTCTCTTTT 900  
CCTGCTTTTC TCTTGAAGA GGAAGGACT CTGTGAGGC CAGGCTGAG TAGATGAGC 960  
TGCAGCTGCG TCATGCGCTT CTTAGAGCG AGAGAGAGT ATGTCATTTT ACTAAGTTCC 1020  
TAAAGAACCA TTATGACAG CAGACTCTT TGCATATCA GAATCTAGG CACATATAGG 1080  
TTATGACTTG CTCAGATATA TGTAGCTCT AGGGGTAAA TCAAGGCATC ACATTTCTTG 1140  
TTGAGGGGCG AGGATAGGC TGTGATTTGC TACACTTTTT TTTTATTAG CATATCTTT 1200  
TTGACTTTT CCTGTGAAG TCGAAGAGC GTACACTTTT CCCAAATGTA GACTAGATC 1260  
TGGAGATGC CAGCACTGT ATAGTTCTGC TTTCGAGAG AGGAGAGCT TTTAGAAGCC 1320  
AATGATCTT AATTTTATT GCTCACTCTT GCTTTTTCG GTTAGAATAT TCACATGAG 1380  
AATGATCTT AAGAGAGAT GCTTGAAGC ATGGGTATC AGGAAGGCT ACCTAAGTTC 1440  
ACATATCTGC AAGCAGACA GCTACCAAGC ATTACTTAGC AGCAGGAAA TGATTTATTT 1500



5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

TGAGTTCCCTG TGTGTCCAAA ACTGAGGCAC CATGTCTCTT GAAGAAGTC CACTCAGAG 1560  
CTGGGGCGGG TGGCTCAC CTTGANTCC AGCAATTTGG GAGGCGGAG GCGGGGGGG 1620  
ATTACACGGG GGT 1634

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1795 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 285:  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 285:

TTCCCCCCCAG GTTGGCTTCC TTGGATTTCT TTCTTGTGTA TCAGGTTTG ATTGAGAGA 60  
CAACCCCTTC TTGTGCACC TCAGTAATCA GCTCACTGTG GAGGAGAGC TGGGGCAGAG 120  
CTGACCTTGA TGTTCATTTT TACCCCCCAA GAGCGCAAAA ACTCTGTGTG GACACAGGAT 180  
GGACCCCTAG CCAGATCTCT CGACAGCTT GTGTCTCTGG CAGCTGAGC CTTGCCCATG 240  
TTAGAGAGAG AGCTCATGTA TCCCCGGGGA CTTGGGGACA TCAGGACAGT GTTCGGGGCG 300  
CCCTTGACA TTACAGAGCT GTGATTTGG CTTATTTCTC GCAATATCCC GCGGACAGCC 360  
AGGCTTTTGG ACTGGCAGAT TGGCTCTTTC TGGCGGGGCC TCTTCAGCA GCGGGGGCCC 420  
TCATCTCTGA TGGCTGTGCT GGTATGAT CTTNCTGAG TCTATCTGAC GAGCTGAGG 480  
GAGGCTTTTG GGGATCTGGC CTTTCTTCTC TATGACAGC ATGTGTGAGA GGTATTTGTT 540  
GTCTCTGGA AGCCACACAG CTTCCAGCGG CAGCCCTTCA AGGCTCTCAG CACAAAGGGG 600  
CGCATGTGTA TGTCTGAGG TGGGGAGCTA GTATTTGTGC CCATTTGTGA AGCATTTCTG 660  
GAGGACTTTG CTTGTCTGGG TGAAGGCTG GTGCAGACTG TGGAGGCGCG AAGTGAAGG 720  
TGAAGCTTTT GATCCAGCT CTGGAGCAG CTGTAGAGCG ACGAGAGGAC ATTGAGCTTC 780  
TAGAGCAGTA TGTCTATAG ATGACTTCCA CCGCTCTTGG ACNTGATTC TCCATGAGG 840  
GCTGTCTGCG TGAATATCT GATCATCTC CAACAAAGC CAGGCCGAG TTCTCTCTTG 900  
ATGCTCTCAG ATTGGGGG GGGGAGTGG GCGCATTTAG TCTCTCTGGC CTCACCTCC 960  
CAGAGAGAGA GTGGAGCCA GCTCAGAGA GGAAGTGAAC CAGAGATTC CATCCACCTA 1020  
TTAGCCCTGG GCTTGAGCT CCGTGGATTT TCCCACTCTT TCTTAGCTCT TCTTCCAGTA 1080  
ACAGAGAGG GATATGTGTC CTGGAGAGG CTCCTCTCTCC TTCTCTCTCC CAGGAGCTGT 1140  
GCTTAGACTT AGCATGCTT TCACTGAGT GTCAAGGCTT TAGATGGGAC CACCGGAAA 1200  
TGTGGCCCTT CTGAGTACA TCAGCGAGC TGAGCATGCG AAGGGGCTA TATGTGTATG 1260

5  
10  
15  
20  
25  
30  
35  
40  
45  
50  
55  
60

ATTAGACAC ATTGAGGAG CAGATGCCC TCTGTGTGTT ATGCCACTTC CGAGGTTGGA 1320  
GACATGTGAA AAGACCGAG GACAGGAAG GATTGGGTAG GTGAGGGGT CAGGGGACTG 1380  
GTACTACCC ANCTTTGAG AGGTCCAAA AGCACTGGG GCTACTCGTT AGCTCATCT 1440  
GCGCTGGCTG TTGCGCCTT CATGTACAA ACTGCCACTA CTATGTACT CGAGTGGGT 1500  
TGCAGAGTG GGGAGACTC AAGTCTTACT CCCCAGAGC TCCCAGGGCC CAGAGGAG 1560  
AATCTGGCT CTTTTCAGTC TGGTCTAGAC CCACCTTTCTG GTAGGCTCTC TGTCTCTCT 1620  
AATCTGGCT GTTTTTCAG ACTAGCTCA AATAGTGGC CTTCTTAGC CCATCCCTCG 1680  
CCCCAGGCT GAGGTGATCT TTCTCTCTC TCACTATTA GAGCATTTAC TGTCTCTCA 1740  
GTCTCTTTG CAGGACACA CAGTGGATA AATCTATTTG TTTTGAATC TGATT 1795

(2) INFORMATION FOR SEQ ID NO: 286:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 858 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 286:

TCTCTTTTC GTCTCTGTC TACTCTGGG CCGCTTCTTC GCGCTCTGTC GGTTCGCA 60  
GCTCTGGAG GAGATCTGG CTTCACTTTC GAGCGGATG AATCCCTCT TGTCTCATTT 120  
TCTGAGCTC TTCCCGCTGA AGGTATTTGG CTACAGCCA GATCCCTTGA ACTACAAAT 180  
AGCTGTGGC TTTCTGGAC TGTGGCTGG GTTCTCTCTG GTCATGGGC CAGCGATCT 240  
GCAAGAGTC AGTAACTTCT TCTTATCT CTTCACTAT CCGAGCAAT GTCTGCTCTG 300  
AGCTCTGAA GATCACTTAA GCACCTGTAT CCGAGCAAT GTCTGCTCTG GGTTCCTCT 360  
GCTCTGAT GTGGGAGC TTTTACCCA GACTTGAAG GTGGTCAAC CCACTAGAA 420  
GAGAGCTTA AGTACATCA AGAATCTG GAAGTAGAG ATCTCTCTCT CTTTATGCA 480  
TCACTCTCT ACAGAGGAA CATGTGAGA CACAGAGTCT ATCATCTTCT TACAGTATA 540  
ATATCAGG TCAGCAGTG TTGAAGAGA CATTTTGTCT ACTTGGACT GCTTCTCTT 600  
TTTACGTTTA CTACTTTT GTAGGAGTA CATGTATGTC ATATTACAT TCTCTATGTC 660  
ATATGAAT ACATATGAG CAGAAAGA ATTTAAATCA ACATATTC TGAATGCCA 720  
AATACACT TTTATGCT TGGTATAGT ATACCTCTGA ACTTTTCT GTGCTTTTA 780  
ACAGATAT ATTTTTTT ATGAAATA AAGCATATA TCTATTTTA TTTCTCTCTT 840

TTAAATACCTT ATAACTA

858

5

(2) INFORMATION FOR SEQ ID NO: 287:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 915 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

10

(1x1) SEQUENCE DESCRIPTION: SEQ ID NO: 287:

GAATTCGGCA CGAGCGGCG CATGGCGCTC CTGCTTTTCG TCTCGCTGT ACTCGTGGC 60  
GGCTCTTTCG GGTCTGTGGG GTTGGCGAG CTCTGGAGG AGATCTGGC TCCAGTTTCG 120  
GAGCGGATGA ATGCGCTGTT GTGACGTTT GCTGAGGTT TCCGCTGAA GTATTTTGGC 180  
TACGAGCCAG ATCGCTGAA ATCCCAATA GCTGTGGCT TTCTGCAACT GCTGCTGGG 240  
TTGCTCTGG TCAATGGGCG ACCGATCTG CAGAGATCA GTAACTTGT TTGATTTCTG 300  
CTCATATGG GGGTATCTT CACTTGGCA GCTCTGAAG ATGACTAAG CACTGTATC 360  
CGAGGCTTG TGTGCTGGG GTTCTCTCG CTGCTGATG TCGGCGAGT CTTAGCGCG 420  
ACTAAGAAG TGTCTAGAC CACTAGGAG AAGACTCTAA GTACATCTAA GGAATCTCG 480  
AAGTAGACA TCTCTGTCT TTTATGCCAT GCACTGTCA CAGCGGAC ATGTAGAAC 540  
ACAGATCTA TCACTGTAT ACCAGTATA TATCGAGGTT CAGCAGTGT TGAAGAGAC 600  
ATTTTCTTA CTGCGAGT CTTCCTCTT TTAGCTTAC TACTCTTTG TGAGAGTAC 660  
ATGTTATGCA TATTAACAT CTGATGTCA TATGAATAA CAAATTAAG AGAAGAGAA 720  
TTTAAATCAA CCAAAATCT GATGCGGCA ATACCACTT TTAATGCTT GGTATAGTA 780  
TACCTCTGAA CTCTTTCTG TGGCTTAAA CAGATATATA TTTTCTTAA ATGAATATA 840  
AACCTATAT CCAATTAT TCTCTCTT TAAACCTTA TAACTATTA TAAATATTA 900  
AAAAAAAA CTGCA 915

50

(2) INFORMATION FOR SEQ ID NO: 288:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1517 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

55

(1x1) SEQUENCE DESCRIPTION: SEQ ID NO: 288:

60

CTTTGTGCA ACTATGAGT CCCCCGGCT GCAGATATTC GGCATGAGT TCTGATCTG 60  
AAGATATCT GATTTCTCT GAGATATCA AAGGCTGCG GAGGAGCC CAAACCCCA 120  
AAGCAGGCC TGAAGAGCC AGGAGAGCC CTCAGGTGAC CAGGCTAGC CTTCAAGAA 180  
CGAGCTTAG CCTCAGAG ACCAGGCGA GCTCTAGAC GCTGAGGCG AGCGATGTA 240  
CGACTCACT TGAAGCTTG CCAAGAGAC AGTCTGACA GACAAGATG GGCACAGTG 300  
GAGCTGAG TCTTCTGCA CAGGAGCAA CAGGGGATT GTCTATGAG CTGACGCCAC 360  
CTCACGCTC ACTGTGACT CAGGACACA GAAGCAAG TTCTCACTCA TACTGATGC 420  
CAGGATGCG GCTTTGTTC ATGAGCAGAA CTTCCTTCAG GGGGCGCCA AGCTCTGCA 480  
AGTCAAGAG TGAAGAGCC TGTACTGAC CCACTGCTG GCAATCCCTA CTTGCAATGG 540  
TTTCGCTGT CACAGACAA ATTAGAGTT CTTCGTGTTA CCAAGCTTGG GAGAGAGCT 600  
TGAGTGGCC CTGATGTCA GCGCAAGCA TGTGCTGTC AGAGAGTCT GTGCTCAGG 660  
TGGCTTGGG GCTGCTGAT GCTCTGAGT TCTTCAATGA GAATGATAT GTTCAATGAA 720  
ATGTGACGC TGAATATAT TTTGTGATC CAGAGACCA GATTCAGATG ACTTTGGCAG 780  
GCTATGCTT GCTTTGCTC TATGCGCAA GTTGGCAACA GGTGGCTTAC GTGGAAGGCA 840  
GAGAGACT TCAAGAGGG GACTTGAAT TTCAATGACA TGAGCTTCCA CAGGATGTC 900  
GGGCGCTCC GCGCGGTGA GCTCAGAGC CTGGGTATAT GCAATCTGAA GTGTTCTTAT 960  
GGTATTCTC CATGACAAA TTGCTTTCCA AATATAGGA CATCATGAG CAATACAGA 1020  
AGTTGCTTG GAATCATAT TAAATGAGC TGAATCTTGT CATGCCAGAA ACAAGGCTTG 1080  
GTCAAGCTCA CGAGTTTCCA GTTTTCCACC TCCGCCAGT TCTTCCGAT GTTTGAACCA 1140  
GATATCTCG TTATCGAGC GCTTCGCGGG GACCACTCC CTGCTTTTGA GTCAAGCACA 1200  
GACAGCTTAC TTGAGCGGCC GCTGGCGCC CACATTCAC TGAATCTGTC GATGTCACA 1260  
GTGACGCTT CTCAGGACA GATGAGCTC CTGAATGCA GCTTCTTTC TTGGAACCTA 1320  
CGAGTTAAA TCTGCTCAA ATGTTTGAAT ACCGCGCAT GCGGCTCAC AGCCAGAGC 1380  
TCCCTGACA GTTGTGCTGT GTGTGTGCT GTGTGTGCT GTGTGTGCT GTGACGGGTG 1440  
GGAGGTCAC CTTTGGGAT GCGGTGTGCC CCGAGAGCT GTAAATATA AATCTTAT 1500  
TTCCAAAAA AAAAAA 1517

(2) INFORMATION FOR SEQ ID NO: 289:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3865 base pairs  
(B) TYPE: nucleic acid

55

60

(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 289:

5 TCGAGGGGGG GAGCTTCTCTT GAGAGGTGGG CCGAGGCTTG GCGCTTCGCA CTTCATCTTC 60  
TGAGCTTCTT CTCTCTCAT TTGGTGCAAT GTCTTTCTTG CAGCTGCTTT TCAGCAGG 120  
TGTTTTCAGT GGGGGGAGCT AGCGTGAAT GACAGGATG GGAAGCGCA GGTTCATTTT 180  
ACTCAGTCTT TCTCTACTCA ATGAGGGGCA CCGAGTGTCT CTAGGGCAGG CTGGTGTG 240  
GTCTCTTAGG TATGAGCTTC TCTTACTGTA CTCTTCGGGA ATGTTTACTT TCTTATTTTC 300  
AGCTGTGACC ACTGTCTTAG GCAAGCTGGC TTGCTCCATG GCGCTGTGCG GTCTCAGCGA 360  
GGTGTGCTGC CCGGAGGGG CAGCGCTTCT TCTTGTATCC TCTTTCCTTA ACAGTGAATT 420  
GGCTTGAAT CTGGCAGGA ACCTTGCTTT TAGCTTCACC ACCAGAGAGA GAGGTGTACA 480  
TGACTTCGC GCGGCTTCAC CAGGCTGGG AACAGAGGGG ATGTGTTGAG AGCGAGTTTC 540  
CTCTGGGCTT CTCTAGGCTG TTTTCTACTA GTCACTACTG TCTTCTCTTT GTAGCTATTC 600  
ATCTCATTTT CTCTGCTTGC CTGTGGGAG TTGGAGATG CTCTGGGTG TAGCTGTGAC 660  
CTGCGACTG AGTTGGGAAA AGAGATAT CAGTGGGAC TGTCTGCTC AGAGCTCTTG 720  
ATCTACCGCA CCGCTTAGGA TCGAGACTG GGTCAAGCT GATGAAAC AGCGCTGGC 780  
AGCAGCTG GAGATGGCT GAGTGGGAG AGAGCTGAC TTCTCTTTC CTCTCTCTCC 840  
TCCAGATTTA CTGAGCTCT ATCTGTTAG GATCTTCTGA GCTTGTTCG CTCTGGGTG 900  
GGACAGGA CAAAGAGAA GGGAGGCTT AGAGAGGCA GCGCTCTTCT GTCTCTGGG 960  
GTAAATGAGC TTGACTTGA GTAAATGAG AGACCAAGG CCTCTGATTT TTATTTTCCA 1020  
TAAATGTTTA GAGTATATA TATACATATA TATATTTCTT TAAATTTTGG AGTCTTTGAT 1080  
ATGTCTAATA ATCTATCC TCTGCTCTGA AGCTGTAGT AGACATGA AGAAGACTGT 1140  
GTTCATTTTA AGATGTTTAA TTAATGATT GAAACTGGC TGTGGCTACT GCTCTTTAAT 1200  
GTGGGGGGA CAGGCAATG/GCTGGGCCC ACNTTNGAA GGGAAATGT TTGTGCTGCT 1260  
GCACATTT GACCCAGTA TGGGCTCTT CTGCTTAGTA CTGCAAGG GACTGTTAG 1320  
GTGTCTTCT CACTCTTAC CCGGACCCC CCHTTAGGG TAAAGGATC CCGAGCTAG 1380  
GTGAGGGGCC AGCACTGGC TCACATGAG TTCTCTCTCT AGATGTCCA GCTCAATCC 1440  
AGACACTTG TTGAGCTTT TTATTTGAT TTATGAGCT GATGGGATTT TGACTGAGAT 1500  
GCTTATGGA GAGTATCCC ACCCTCATG AGACAGAT CACTCTCTGC CATCTATCT 1560  
GCTGTACTT ATGACAGC AGCTATTTA GGGATGTCC CAGCTTACT GATCAGGG 1620  
AAATTCAGG AGCGCTGGGG CAGGCGCTGG NCCGAGTGC CAGGCTCAG AGTAAAGGA 1680

CATTGGAAA GTTCCAAAC ACTTGGTAGA CCACTAGGTT CTCTGTTTTC CCTTCCCTTT 1740  
CCTTTTCAAA TCCGACAGTT TCTGTGTGGG GAGAGCTCTT AATTAGCTTA GTCCAGTAC 1800  
CAGATCCGAG CTAGGGGGGC AGCTGNTTGG GATTAAGTCCA AGAAAACTGG GGCACAGTA 1860  
TTTTTCCAT TATTAAGATCT GTGCAATAAA TTTTAAATG AGTTATATGG AAACAGATT 1920  
TCTCCAGCTG CAGGGGAG AGGTTAGGGC TGAGTCTCTT GCTGTGGGCC AGCGCTTTT 1980  
AGGGTTGCTT CTCTACTGC ACCGAGCAG GATGATCTTG GTTCTTGGGG AGGTTAGCTT 2040  
GCGCTTGGC GAGTTCTGCA CCGAATAAG ACTCCAAAC GGTGTCTTTC GTGTCTGAG 2100  
AGATGGTTAA ATGGGTGATG GATGAGCAG ACTGAGAGGA CAGCAGATGA CTGATGTGTC 2160  
GAGAGAGGGG GAGATGCTT GGGCTGGCTA GCTAATGTTTC CCGCTTTTCA GCGATTTACA 2220  
GGAATTCGAG CCGAGCTTGG TCTGAGATT GTTTCCTTC CACTGTGGGA TGCATCTCTC 2280  
AGAAATTTTG AGTCAAGCTT GCACTTCTC GAGACTTTTC TTCTTGGGCT TGAGCTCTTC 2340  
ATCTGTTTGG CCGTTTCTAT AGCGTTTAC AAGAGCTGC TCACAGGAG CAGAGCTTGC 2400  
CGAGAGTTC AGAGTTTCAA CTGGCGTTT ATCGCTTTTA TAGAGCACA CAGAGCATG 2460  
CCTTGGGACT GCACTCTCTT CACTTCTTGG GTTTCAGGT TCACAGCAG CACTAGCAGC 2520  
CTGTCTGCA GTTCTCTCTT GGGGAGAAC TGTACAGGC CCGTCAAGC AGTCCGTGCT 2580  
TGAGTTTGG CAGGTCTAT CTCTCTTACA TAGAGGCTGT CTGATCTTGG GTCTCTTCTC 2640  
ACAGTATGA TTTTCCCCAC AGCGATATCC AGCGGGGATG GATGAGCTC CTCTGTCTT 2700  
GATTTCTTGG CAGCGTTTGG GCGATTTGCT TCTGCTTTGA GCGATCTTGG TAGGCGAGC 2760  
TGGCAGTTT TTTCAGGGCA GGGGTATTA ACTTTTCCCG GATTTGATCC AGCAACTTGT 2820  
TCAGTGGCAC TTTCACAGAA TTCTTCAGGT CTCCAGGATG TACAGCTTCA CAGACAAAT 2880  
CCTTTTCCAG GTCAAGTTAA GCTGTGTAG TTTTTGTTC ACCCGATTTT TCATCTGTA 2940  
GATTCAGAA CTGGACTTA AGGGAAAA GCAATCTTT GATGAGGAG AGAACCCAT 3000  
TGTCTTCAC ATTTCTGGC TCACAGAGG CTTTCTTAG TTTTCTTCT ACATCTCTCT 3060  
TCCGATCAG GAGATATATC TTGGACTCTT CTCTGAGGA GCTCATTTTG CTGCTGTATA 3120  
ATCTGAGAC CATAAGATTC ATAGATGGA CCGTTTTTGA ATAGCGAAT CAGAGGAGCT 3180  
ACTTCTTGC AAAGGTGAAA ATCTTCTCT GATCATATCC TCCAAATTTG GCATCTACTT 3240  
TTAAATATC TTCAATCAA GCGTGCAGTC CCGGTATTA GAGGCGACTC AGCAAAAGCT 3300  
GCTCAGCTG CTTHACAGC TCAGCTCAG CTTTCTTGA ATCTGTCTGT GTGACAGG 3360  
AGAGAGTCT GTACAGATCT AGTGTGACT CTTTGCTAG CTGGTAAATCA GTGCTTTGA 3420  
TGAGCTTGA CTCTTCAG GGCAGGCA TGTCTCTCAG GATTCCTTGG ATACATTTCT 3480

5 CATTGACCT GACTGAGT TCTAGAGT CCATGAGG TTTCAATTTA TCCAGATAG 3540  
CTGAGAGT CCACACAG ATTCTTACT CACACCTG CTTTAGAG TCTGATCT 3600  
TTGACATGG CACAACTA GGCATATG GTTTGCGGT GTTTCGCTT CCGCATTA 3660  
TTTAAATTC CCGCTCTTC AGTATCTCT TCACTTCTC TTGCGCCAG ACCTCTGCA 3720  
10 GTTTCGCTT GATAGATG AGTTCTCTT CAGGCTGAG AGGCTGCGG ATGCTGCGT 3780  
ACCGCTGCTT CCGCGCTCA GCGCGCAC AGAGCGCTT CTTGCTGAC CTGCGCGC 3840  
15 GCTGCGCGG AACTGCTAG CGAT 3865

## (2) INFORMATION FOR SEQ ID NO: 290:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1910 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 290:

30 AGGAGAGA GAGAGGGG TGTGCGGG CCGCTTACC AGAGCGAG GAGCGGAG 60  
AGGAGTGG CTGTGCGG GCGAGCGG GAGCGAGG CCGCGCGG CCGGAGGAC 120  
GGCTTACA GCGAGCGG CTCTAGAG GAGGAGAG GCGAGGAG CTCAGGCTC 180  
35 ACCAGCGG CTGCGCGG CCGCGACT CCGAGGCTT TCGGAGCTT GCGGCTCTC 240  
CTGAGGCTT CCGAGGAG CCGCTTACA GAGCGCAT GATGCTCTT GTTCTCTC 300  
40 TCGAGCTG TGGCGCGG GTTCTGCG AGTCACTG AAGAGAGA GGAATGAG 360  
CTTTTCAAT ATGATACA GAGCGTGG ATGCGGAG TGTGTTTC TGTGCTCTC 420  
TTCTGCTT GATCTCTT TATCTTACT CCGAGGAG AGTCACTT CATTAGAG 480  
45 CCGCGCGG CAGAGAGA GAGCGCGG GTGAGAGG TCTATGAG CATTGACA 540  
GAGCGCGG AAGAGAGA CTGATGAG CCGATGAG GAGGCTCT GAGAGTGG 600  
50 GCGCTCTT GAGCTTGG ATGATGAG CTATGCTTA GAGAGCGG CACTTACA 660  
ACGCGCTT CCGAGAGA AGCGAGAG TGTGCTCT CCGAGCTT CCGCTTAC 720  
ACATCTCT CACTGAGA TCGATTAC ACTTCTCT CACTGAGG CTGCGCTCT 780  
55 GCGAGCTG CTGATGCT GTTGTGCT GTTGTGCT GACTGTGT GTTGTGAG 840  
TGTGCTCT GTGCTACT GTTGTGCT GATGCTGT TGTGCTGT ACTGTGACT 900  
60 CCGTTTCCA GCGAGGCT GAGCGAGG CCGATGAG CCGCTGCT CCGCTGCG 960

TCATCACT TCTGCTCT TGTGCGAG TGTGCGAG GAGCGCGG CTGCGCTCT 1020  
TTAGGAGT GTAGGAGG AGCAGGGA GTGCTCTT GTGCTCTT GTGCTCTT 1080  
5 GTTGTGAG ACTTGTCT CATTCTCT GAGCTCTT CACTCTCT ACATCACT 1140  
TCTTCTTA TCCACTGA TCCAGCTG AGGCTCTT AGCAGCTGA GATCAAGG 1200  
10 AAGAGCTG TGAAGGAG GTTGAAGT GCGAGCTAT GCGCTCTCT GTTAAATTC 1260  
TTCCAGGG CTTCAGAG GATGCGCT CTGCGCGG GTTCACTA GCGCGCGG 1320  
ATTCAAGG CAGGCTCT ACTTCTCT TCGGAGCT GTGCGCTGA TATCTCTCA 1380  
15 GCAATCACT CATGCGCT GCGAGCTAC CCGTCTGAG CTTCTCTCT TCTGAGCTT 1440  
CAATCAAG CCGAGCTAT CCGATGAG ACTAGAGT CTGATGCT GTGCTCTGA 1500  
20 GCAATGAT GAGAGCTCT TGTGCTGG GCGCGAGG ACCGAGAG ATGAGGAG 1560  
AGCAGGCG TTTCTCTC TCGTACTC CCGTACTG GCGAGAGG CCACTCTCG 1620  
CATCTTCT TCTGCTCT GTGCTGAG CCGTACTG AGTGTGCT GAGACTGAG 1680  
25 AGCTCTCT CAGCGCTG GAGAGTGG AGTGTGAG TCTATGAG AGTGTGACT 1740  
CAACCAAT CCGCGCTC CTGCTCTG TGTGCGCG GAGCGAGG AAGCTGTG 1800  
30 CTGAGAGA TTTCTCTT CTGATGCT ATCTCTCT AAGAGAGA GAGAGAGA 1860  
ATTAATAT CTTTCTCT TAAAGAAA AAGAGAGG AGGCGCGG 1910

## (2) INFORMATION FOR SEQ ID NO: 291:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3776 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 291:

45 GCGAGCTG TTTGAGCT CCGTCTCT GCGCTGCA CTGCGCTG CAGCTGCG 60  
ATCAGAGG AGCTTCTT CCGCGAGG TGTGCGCG TCGAGAGG CAGAGAGG 120  
50 GCGCTGCG GTTGTCTG CCGTCTG CCGCGCTG GAGAGGAG GCGCGCGG 180  
AGTGTGCT TTTGAGCT TGTGAGCT TACCGCGG CCGCGAGG CCGCGCGG 240  
300 GAGAGGCG CCGCTGCT TGTGAGCT ATTCAGAG TGTGAGCT TCAAGAGG 360  
420 AAGTGTCT GTTCTGCT CAGCGAGG TGTGAGCT GAGTGTCT CAGCTGAG 480  
60 CAGTGTCT GAGAGAGG AAGAGAGG AGCTGTCT CAGCTGCT TGTGAGTA

540 CTTCTCTCTGA TCGAGAGATT CGGATGAATA GAGTTCTTTC GAAATACCTT COTCTAGCC  
5 TAGGGAGATT CATTGACATC CAGCCATGCC CTGATCTGAA GTACGGGAAA CTTATCTCATG  
600 TCGCTGCCAT TGAATGACAA GTGAGAGGA TTAATCTGTA TCTCTTGAG GTATACCTTA  
660 AGCCCTACTT CCGTGAGAGG TATCTGACCA TCCGGAGAGG AGACATTTT CTTCTCTCTG  
720 GTGGATGCG TCGCTGTGAG TTCAATGTGG TCGAACAGA TCTTAGCCCT TATTTGCATTG  
780 TTCTCTCAGA CACATGTATC CACTGTGAG GGGAGCTTAT CAACAGAGG GATGAGGAG  
840 AGTCTCTGAA TGAATGTGG TATGATGACA TTGTGTGCTG CAGGAGCG CTAGCTCAGA  
900 TAAAGAGATT GTTGAGACTG CCGCTGAGAC ATCTGCGCT CTTTAAAGCA ATTGTGTGTA  
960 AGCTCTTAG AGGATCTCG CTCTAGGAC CTCTGGGAC AGGAGAGAC CTGATCTCTC  
1020 GAGCTGTAGC AATGAGACT GAGGCTTCT TCTCTTTGAT CAATGTCTCT GAGATCTGTA  
1080 GGAATTTGCG TGTGTGCTCT GAGAGCAAC TTCTTAAAGC CTTTGAAGAG CCGTAGAGTA  
1140 ATGCTCTCTG CATTATCTTC ATGTATGAGC TAGATGCCAT CGCTTCCAAA AGAGAGAAAA  
1200 CTCAATGCGA GTTGAGCGG CCAATGTAT CAGATTTTAT GAGCTCTATG GATGCGCTAA  
1260 AGCGAGGGC AGATGTGAT GTTATGTGAG CAGCAACAG ACCCAAGAGC ATGTAGCCAG  
1320 CTCTACGGCG ATTGTGTGCG TTGATAGCG AGGTAGATAT TCGAATTTCT GATCTAGAG  
1380 GAGCTCTTGA GATCTCTGAG ATCTATGCA AGACATGAA CTTGTGCGAT GATGTGAGC  
1440 TCGAACATTA GCGATGAGA CTGACGGCA TGTGTGTGCT GACTTAGAGC CCGTGTGCTC  
1500 AGAGGCTGCT CTGAGAGCA TGTGAGGAA GATGTATCTC ATGTACTTAG AGATGTGAGC  
1560 CATTGATGCC GAGTCTATGA ACTCTCTAGC AATTACTATG GATGACTTTC GTTGGGCTTT  
1620 GAGCGAGAT TACCGATCG CACTGTGGA ACCGTGTGTA GAGTGTGAC AGGTAACTTG  
1680 GGAAGACATC GGGGCGCTAG AGATGTGCA ACCTGACTTA CAGGAGCTG TCTAGTATCC  
1740 TGTGTGAGC CAGACATAT TCTGTAGTT TGGATGACA CTTTCCAGG GATTTCTGTT  
1800 CTATGAGACT CCGGCTGTG GGAATCTTT GTTGGGAAA GCGATGTCTA ATGATGCCA  
1860 GCGCACTTC ATCTGCTGTA AGGTGCTGTA CTTCTGACC ATGTGTGTTG GGAATCTGTA  
1920 GCGCAATGTC AGAATATCT TTGACAGCG CCGCAAGCT GCGGCTGTG TCTATTTCTT  
1980 TGTGTAGCTG GATGTGATG CGAGGCTG TCGAGGTATC ATTTGAGATG GTGTGTGGGC  
2040 TGTGTAGCGA GTATCTACC AGATCTGAC AGAATGTGAT GCGATGTCTA CAAAAAATA  
2100 TGTGTATC ATTTGGCTTA CAAAGCGGCC TGAATATAT GATCTGTGCA TCTCTAGAC  
2160 TGGCGTCTT GATGACTCA TGTACATCC ACTTCTGAT GAGAGTCCC GTTGTGCCAT  
2220 CCGTAGGCTT AACTGTGGA AGTCCCAAT TCGCAAGAT GTGAGCTTGG AGTTCTTGGC  
2280

2340 TAAATGACT AATGGCTTCT CTGAGCTGA CCGACAGAG ATTTCGAGC GTCTTTCGA  
2400 CCGCGCATC CCGTGAATCCA TCGAGATGA GATTAGGCGA GACAGAGGA GCGAGCATA  
2460 CCGATAGCC ATGAGGTAG AGAGATGTA TCCAGTGCCT GAGATCCCTC GAGATCACTT  
2520 TCGAGAGCC ATCGCTTTTC CCGCCCTTTC TGTCAATGAC ATGACATTC GGAATATGA  
2580 GATTTTGGC CAGACCTTTC AGCAGATCG GGGCTTTGGC AGCTTCAGAT TCCCTTCAGG  
2640 GAGCGGCT GAGCTGCGC CAGTCAAGG CAGTCAAGGC GCGACAGGTG CAGTGTATTA  
2700 CACAGAGAC AATGATGATG ACCTGTATG CTAATGTGTG GTGGCGAGG TCGATGTAGC  
2760 TGGCTGCTT GAGCTTCTT CCGTGGGCTT GGGGCGCTT CCGCAGAGGA GCGAGCAGG  
2820 GTGCGCCAC AGCTGTCTC ATTCTCACT CTGACAGATT CAGCTACAGT CTGACTCTG  
2880 ACAGGGGCTT TCTGTTCGA AATATGAAA CAAAGCGAT AATATTAAG CCAATTTCTAT  
2940 TTGTAGGCG CAGATGAT TACACAGG GAATTTGGCC TTGGCTATG CCAATTTCTT  
3000 TGTATTTGG GCGATGAG GCGACTGTG TGGGTGTGTA ACCAGGAC TACTGCGAC  
3060 TCGCAGATA AGCATCTOC ACTTGACTA ATGCTGTGCG AGCGCTCTCT TCGCCTATC  
3120 CAGCTGCTT AGTGGGTAG GGGCAGAT TCTGTGAT TTAATAGAG ACTTGTGTA  
3180 TTTATTTTAC ATCTTTTGA GTTATTTTG GAATACTAT CACAGCAAT TCTTAACCA  
3240 AAAAAATGA TTTTGTAAA GGCATATA CTTTGTGCT AATGTGGTAA AAAAAAATA  
3276 AAAAAAGGG GCGGCTCTTA AAGNCCANN CTTCTT

## (2) INFORMATION FOR SEQ ID NO: 292:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1695 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 292:

50 TTGCAATGCT TGAATTCCTT TCTCAGGCC AGCTTAGAG AAGAGTTGG TACTTCCAGA 60  
GTTGAGCAG CAGCGCGCG TTCTCGCGG GTGAGGCGG AGCTGAAGT ACAGCGAGG 120  
CGAGAGCAC GTTGTGTGG GCGAGAGAG GGGCTGTGCC CAGGAGAGAG AGAATACCTT 180  
TCTGAGAAA CAGCAACAG CTGACTGCT GTGACAGAG GGAACAGAT GCGCGCGCG 240  
AAGGAGGCTT CTGGTGTAGG ACCCACTGG GCGTCCCGCC GCTGTCTCTG CTGACATGCG 300  
CGTTGCGCG AGGTTGCGG ACCGCTTGG CTAAGCATT TGAATCGCTT TTGGGTGATA 360

5 CCGCGTCTTG CCACTCGGCGC TGTGAGTTGA CTTACCGCTT GCACACCTAC CTTAGCGAG 420  
AGAGTGTGTA GCGATGCTCAG AGAGTTGCA GCGTGTGTTTC AATTGTGACG TTGTGCGATG 480  
10 ATGGGATTTGA CTTAAATGGA ACTAAATGCG AATGTGAATC TGCATGTACA GAGCATATTT 540  
CCCAATCTGA TCGACATATG GCTTCCCATC TTGGTTGCGA GAATCAGCTG CCAATCGCTG 600  
AAGTGAGCA AGAAGCACTT AATGCCCTGA TGGCAAAAT GCACCTACTC TTTCCTCTTA 660  
CTCTGTGCGG GTCAATCTGCG AGTGACATCA TGGACTCGCG ACAGAGCTTC ATAACTCTTT 720  
CATGAGCTTT TTATCTTCAA GCGATGACG GAATAATAGT TATATTCGAG TCTAGCGCAG 780  
15 AATTCAGTGA GCGACCAAT TTGAGCGCG AGGCTACAAA TTTCAGAGGA TCACTCTTAA 840  
GGAATATGTC CTATCTGCA ATOGAAAT CACAGGCGA CAGGAATTTT CTTGAAAGTG 900  
GAGAAAGTGA TGGCTTTTAA AGATCCCTCT CTTTAACTC TGGTGGAT TTAACTACAA 960  
20 CTCTGTCTCT CTGCGTGATG GTATTCCTTT GCAATTTGTTG TGCACCTGTT GCTACAGCTG 1020  
TGGAGCAGTA TTTTCCCTCT GAGAGCTGA GTATCTATG TGACTTTGAG TTATGCAATG 1080  
25 AACAAAGCT AACAGATAT CAGCTTCTTT CTCTGTGCT TGTAGATCT AAACCTGAG 1140  
ATCATGAGA AGCGGCGCT CTAACTACA AAGTGAATCT TGGTCAATCT GAATTTTAC 1200  
CATTTTCTTT TTAAAGACA AGTGAATAG ACATCAAAA TTGCACTCTT CAGAGAGCTT 1260  
30 TTAAATGCTT TTCAATGCTT ATAGGCTTA AGAATCACT ATAAATGCA AATTAAGTTA 1320  
CTCAAACTG TGAAGACTGT AATGCTATA ACTTTATGCG TATTTGTTTT GTATTAATTT 1380  
35 AAGAGTGGA TTTTGGGAT TGTATTATTA TTTTACTAAT ATCTGTAGCT ATTTTGTGTT 1440  
TTGCTTTGCT TATTTTTTTT TTGCTTTTC TTAGCTATGA GCTGATCATT GCTGCTCTTC 1500  
40 ACTCTGCGC ATGATACTGT CAGTTACTT AGTTACAG AGTAATATTT AGTAGAATG 1560  
ATGCTTCTGC TCGAGATGCG CCGACAATC TGTAAATTTG AATTTAGCAG GAATGACCT 1620  
TTATGACAC TACATTTTCA GCACTGAA TCAATTAAT TTATTTGAA TATTTAAAA 1680  
45 AAAAAAAAAA AANCT 1695

50 (2) INFORMATION FOR SEQ ID NO: 293:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1501 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 293:

60 CACTTTCGAC AGTCTTTGCT TCTTTTGTCT TCACTCTCAA ATAGCCCGAG GAGTGGGCTT 60

5 TATCTCTCAA TATCGACAT CTAAGCTTC TCTTGCGGGA TGGGAATTTG GATGCGCAGA 120  
ATCTGTTTTG GATCTCGGCG TTATTTTCAG TGGGTGAAA ACACAGACTG GCGCTTTGCG 180  
20 TCTCTTAATC CTGAGGCTGCG GTAAAGAGA CTTATCTTAC ACCTGTCTTT CCGTACCTTC 240  
TCTTTTGTGA GCGAGGCTTC ATTCATAGTT CCTCAAGAGA GTCTTTGGCT TAAAGCTGTA 300  
10 GCAAGGCTGT GCTAGGTGCG GATTTTGGAG CAATAAGCTC GATTAAGCAT GATTAAGCTTA 360  
TGGAGTGGCG CTGCATAATC AGACAGAAAT GCGTTGAGAA GCGCGAGGCG AGCATGCTCG 420  
TCTCTCAATG ATAGATGATG GAGAGGAGCT CCGTACCTTG GAATAATGGA ATTGAAGCGG 480  
15 TTATCGACAA ATAGAGTGGC TATGTGAGGA TTTTCCGAAA GTTTTGTCCA ATCTTATCAT 540  
TATTAATATT TATTAAGCCAC AGAGACAAAC CAGAAACGGA ATAAATGTTAC TTTCGATGCT 600  
20 TTATTTTTTT GTTCTAGGCG TGGCTTTGTA CATTGAGAG AATGCTATAT GCTGCGCATT 660  
TTGCGCTTAA AGCTCTTACA CTTTCCGCAAT TTATAGCTTA TGGGAAGATA CAGATGTGCA 720  
AGTCTGCTTT TTGTTTTTTT GTTATTAATTT TTTTTTTTTT GCTCTGTGTT ATGACATATT 780  
25 TCGACATGCG ACAGAGTGG AGAGATGCT CTTTGGAGCC TATGTGATG TCACTAGCT 840  
GCACTACTTA TCACTATGCG TCACTGCTGT TTCACTGTA TCTCTCTCTT TTCACTGTA 900  
30 TTGTTATGCG AAATTCGAG ACATATGCG AATGCAAGCG TCACTACTTT GCGAGATGCG 960  
TATCTCTCTT TGAATGATAT AGTATGCGG TGCATATGCA TATCAATC AGCTCTGCTT 1020  
35 TTGCTTTTAA ATGTTAACTA ATGAATGTC AGAGATGCGC CTTAGAAATG TATTTTAAAG 1080  
ATTAAACAGG AGTCTCAAAA AGAATGAGA GCGATGCTTC CTTTNGCCTT GCACTCTACA 1140  
AAGAGAGAG AGACTGTCTT GTTGTAAAC TTTTCAAAA ATCTGATATAT GGTAAAGTAC 1200  
40 TTGAGAGCCT TCAAGAGAT GTCAATCTTT TTGCTGTGCT ALCATGAAA CTTGTGTGAC 1260  
CATTAAGCAT GTTATCAGCT TGTACTGCTC TCAATAGCTT GGGTTTGGAA GAAATATTTG 1320  
GAAATTTTGG CTGTGTTCTG TGAATATAC CTCGCCAAA TATTAATGTA CTGTTGTTTC 1380  
45 TACTTGTGTA TTTCAGACCC TTTTAATAC GCAATTAATTT CTGTGTTCTT AAACAGTATA 1440  
AATAGTTGTA AGTTTGGCTG CATGATGGA AATTAATAC CTGTATCTCT GTTAAAGAAA 1500  
50 A 1501

55 (2) INFORMATION FOR SEQ ID NO: 294:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2683 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

60

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 294:

5 TGATGTGGT CCGGGGTGCG GATTGGCAGC GCGCTGGCCG CCGCTGGTGG TTGTCCGGCC 60  
ATGGCACTGT CCGCGGGGGCT GCGCCGGGAG CTGGCTGAGG CGGTGGCCCG GGGCCGGGTR 120  
CTGGTGGTGG GGGCGGGCGG CATGGCGTGC GAGCTCTCTCA AGAATCTGGT CTTCAACGGT 180  
TTCTCCGACA TCGACCTGAT TGATCTGGAT ACTATTGATG TACGCAACCT CAGACGACAG 240  
TTTTTGTTC AAGAGAAACA TGTGGAGAA TCGAAGCCAC AGTTTGCCAA GGAAGTGTAA 300  
CTGAGTTT ACCGAAAGC TAAATCTTT GCTTACCATG ACAGCATCAT GAACCTGAC 360  
TATATATGAG AATTTTTCG ACGATTATTA CTGGTTATGA ATGCTTTAGA TACAGAGCT 420  
GCGCGAACC ATTATTATAG AATGGCTG GCGAGTGATG TTCTCTTTAT TGAAGTGGCA 480  
ACAGCTGGGT ATCTTGGACA AGTACTACT ATCMAAAGG GTGTGACCGA GTTTATGAG 540  
TGTATCTTCA AGCGGACCA GAGAACCTTT CTTGGCTGTG CAATCTGTAA CACACTTTCA 600  
GAACCTATAC ATTGCATCTG TTGGCAAG TACTTGTCCA ACAGTTGTT TTGGGAGAAA 660  
GATCTGATC AAGAGTATC TCGTGMAGA GCTGACCTG AGCTGCTGTG GGAACGAGG 720  
GAAGCGGAG CCGAGCTAG AGCATCAT GAGATGCTG ACATTAAAGG TATTTACTACT 780  
AAGGAATGGG CTAAATCAG TGGATGAT CCAATTAAAA CTTTATTACA AGCTTTTAA 840  
AGATGATC AGTATCTGT TGACATGGA CAATCATGG CGGAAGAAGA AACCTCCAT 900  
TCTGTGTGAC TGGCTGAGG TGAAGTCA AGGAGAGAA AGCATGCTAT CAGTCAACA 960  
GAATGAACC CAGTTAGGCC TGAAGACCA GCAAGTTCTA GATGTAAAGA GCTATGCCAG 1020  
TCTTTTTC AAGAGCATG AGACTTTGAG AGTTCAATTG CAGAAAGAG GCGATGGAGC 1080  
TGAGCTGATA TGGGATAGG ATGACCATC TCGAATGGAT TTGTGCACT CTGCTGCAAA 1140  
CCTCAGGATG CATTTTTC ATATGATAT GAGAGTAGA TTGTGATACA ATTCATGGCC 1200  
AGGGAATTT ATTCTGTCTA TTGCTACTAC TAATGCCATG ATTCTGGGT TGATATAT 1260  
GGAAGGATG AGATTTTAT CAGGAAAT AGACAGTGC AGAATCAT TTATTGATTA 1320  
ACAGCAAC CAGAAAGA AGCTTCTGT GCTTTGTGCA CTGGATCTC CCAAGCCCAA 1380  
TTTGTATGTA TGTGCGAGA AGCGAGGT GACTGTGGG CTGAATGTCC ATTAAGTGAC 1440  
TCTTCTACC TTACAGACA AGATATGAA AGAATAATTT GCTATGATG CACGATGT 1500  
CGAATTTGAA GATGGAGAG GAAATCTT AATATCTCC GAAGGGGAG AGAGGAGGC 1560  
TAAATATC AAGATTTGT CAGATTGG AATTAGAAAT GCGAGCGCGC TTCAAGGAGA 1620  
TGACTTCTC CAGGACTATA CTTTATGAT CAAATCTTT CATATGAGG ACCTAGAAA 1680

GAACCTTGA TTGAGATTG TTGTGATGC CCGGAAAAA GTGGGGSCA AACAGCTGA 1740  
AGATCTGCC AAGACATA CCATGGCA GTATGATGG AGCTGAGCC TCCACTCCA 1800  
5 CAGCTCAGA CGAGATGAC GTTCTCATAG TTGATTCGGA TGAAGAGAT TCTTCAAATA 1860  
ATCCGAGCT CATGAGAG AGAGAGCG CAGAGGAAA TTAGATGAGA AAGGATCT 1920  
CAGTCMAAG AGTCACTTA TAGACAGAA GGAAGAGCTT GATGATGTCA TAGCATAGA 1980  
TTGACAGAA ATGCTCTTAA ACAGAGCT CTACTATTT AGTTATCTG GCGAGACCA 2040  
GATTTTATG TCTTTTGT CAAAGGAAA AATGTGACG CAGTCACTG AATATGATC 2100  
15 TCTCTCTTT GAAGCATTC ATTTTCTAG AACTGTAGA CAGTTTCCAG TATGCTGTAT 2160  
TGAAGTAGG AATATATTT TAAAAAGCT TTGACMAAG TGTGTGTATA ACCATGTAG 2220  
AGATAAACA ACACATGCA TGTGTCTTT TTATGTAAA TACCTTAGG TATCATTAAT 2280  
20 AGTTTCAAA TATGTGTTT TAGTAAAGTT GATACCTGTT TATTAATATAT ATGCTTTAT 2340  
TTTGTGCTAG AGAGGATTT ATTTTATGCC TAGATCTAAC CATTTTGATA CTCTTACTG 2400  
25 ATTGAACAG ATTCAAGAA GTATGAGTG CTATGCTATG AACTGTGTT TTAAATGTAA 2460  
CATGCACTA TGTATTTAA TGTAAAGCA TTTTAAATTA CTCAGTTTTT CAGTTTGTAC 2520  
CGCTGTGAT GTCTGTAAA GAAGCAAT TTGTGTATAT GTTACAGATT CAGTTTAT 2580  
30 ATATGATG TTTTGTAAA CTCMAATAC GACTATACT ATGACACAAA TAAATGGCA 2640  
TGCTTTCTG TTAATAAAN NACAGAAAA AAAAAAACA AGA 2683

(2) INFORMATION FOR SEQ ID NO: 295:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1454 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 295:

50 GCACTGGGG TGCTCTTAG GCGAGGAGT AGGCTGGGG AGCGCGGCC TGTGGCTG 60  
ACCAAGCTT TCTGTTCAG GTTCAAGCC GATCAGGTG GTCACTGCT TGAAGCGGA 120  
CAGTACTG AGCACTGCT TCTTCAGCA CTTGATGTC GTGCTGCTT CTCTGGAAG 180  
CAGCGCTCG CCGAGAGCTG TTACAGAA CTTCTACTC GAGTTTGGGA ACAGACAC 240  
AGGAGAGATG GAGACTAG AGCTATCA CTTTATGCG GTCAATTTTA CTTACCCAG 300  
TGAAGGAG ATTTGGAGC TGAGTTTAC TGTGGCCAA AAGATGGCTG AGCAGAGAA 360  
60 GCGCCAGCC CTCAGATCC TGTGTACT GCAAGCTTC CAGTGGGA TGCAGCCCC 420

480 TGGTCTCTGC AGGGGCCCC TGGCCGCCAA GACACTCTCG CTCACAGCT CCGAGCTTT  
5 CCTCTTGAT GAGAGCTGTC TCCACTACCC ACTGCCCCGAG TTTCGCCAAG AGCCGCCCA  
600 GAGAGCAGCG TACCGGCTCG AGAGTGGCG CCGGCTTCGG GACCTGGACC GATCTCTAT  
660 GGGCTACAG ACTTACCGC AGCCCTACC CTGGTTTTCG ATGACCTGCA AGCTCATGAC  
720 CTATGCGCA GTGTACCTT GAGCAGCTT GGGGAGGTGC CAGGTGGCCC GCTTAGAGCC  
780 AGCCAGGCC GTGAGTCCA GTGGAGTGG TTATGCCCA GTCTGAGAG CAGAGAGAG  
840 CTATCTGC TGTTCGCTCG CAGTGGAG GCGCTGTGTC GCTTAGAGCT GCTGTGAGC  
900 TACCGGCTA GCGCAGGCA CAGCCAGCT GTCTGTGCA GCTTGAGCC TACTGGGCA  
960 GGGAGCAGG CTTTTGTCT CAGTAAAT GTTTATCTCT CCTTTGGTA CTTTATTTG  
1020 ACTGTCTCG CAGAAATGTC AGATGTGTC TGTGTGTCT TATTTCTTTC TCAATTTGG  
1080 AGTGAGATG CCGGGCCCT CAGGCTGTT CCGTGTGTC TCACTCTCC AGAGTGTGA  
1140 CAGCTGTCA CACAGTGTG GTGTCTGTC TTGTGGGACC GTTTTAAACA GTTGAGCTG  
1200 TGGTCTGAC TTATTTCTCT AGATCTCT TTCTGAAATG TCGATGTCAG TCTTTTGTG  
1260 CTGTGCTGT TGTGTGCT GTTGTCTGTC GCATCTTCT CTAAATCTG AGCTGTGAG  
1320 CAGATGAC ATGTGAGCT CCGACCCAT ATTGTCTTC AAGTGTGAG TCTCCCTGA  
1380 TCGACAGAG TGGAGAGCC GTTGGGCA GGGAGCTCG AGCTGCCAGC AGCAGCTG  
1440 ATTCTGCTG CTTGTATCT CTATTCAT AAGCGAGCT TTGACAGCTG TAAAAAATA  
1454 AAAAAAAAAA AAGN

(2) INFORMATION FOR SEQ ID NO: 296:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 828 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 296:

60 ACCCTGGCAT GCGCCAGAAA CAGATGACCA CCGAGCTTAC ACAGGCTTAC ACTCTCTCA  
120 ATGAGAGAA ATCATTCAC ACTGAGCAG ACATCTATAT GATCTTTTAA GGAAGTGTTT  
180 CCTTATGTC TTGAGAGTAA TATCTGGTAA ACTCTTAAT OCCATGAT ATCTCTAGGC  
240 TGGACAGAA TGGCTGCA TTAGCGAGT AAGAGCTCA GTCCAGTAA ATGATCTAT  
300 AGACTCTCT AGCAGCACT ACATGACA CTATGTGAG AGCTGACG CCGCAAGTA

360 GAAGATGTC ATATATCTG CTCTGTGTA GCTCAGAGA CAATTCCAC ACAGACTTA  
420 CAGTTACCC TGAAGTCAG CTCGAGTAA TACGAGAC ATCCAGAAA ATACTTATG  
480 AGGGGGGAG GCTGAGCTG GCGCTTGAG GATGANTGA ATTTGATAG AGATGAGCA  
540 AGCAGAGGG TCTCGAGTG AGAGAGCT GAAATAGAG CAGGGGCTG GATCAGTGG  
600 GTTATTTAG AGCACTTTC CAGATGACC ATGATGCTC ACAGTCTCT GCGTATGCT  
660 GCGAGTGT CCGAGCGCA TGTGTGCCC CAGCCATGT CATTATCAT GTCTCTCAT  
720 GCGACCTCA AAGGTACTT CTCTGTAAA GCTTTGCTT GATATCAGA ATCAAAATTA  
780 ATCAGGATC TTTTCACT GTGTTTTTT CTCTTTGCT CTTCTATCA CTAAAGCTA  
828 TCTATTCAG CTTAGAGCA TAACTATTA TTTTGTTC TCACTACA

(2) INFORMATION FOR SEQ ID NO: 297:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2416 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 297:

60 TCAATTTCA TTAATGAGA TGAAGCTG TAAATGACA TTGTGAGG TCTCAGTTT  
120 ACGAAACT ACTATGACA TCAATCTTCA ACAGCAGAG TCTGATTTCA GCGAATTT  
180 TTTTCTTT GAGAAAGAG TGAAGTGGG CAGACTTTT AGTCTGAGG GAGCTAGTGG  
240 AATCTAGAC AATAGAGTC ATCGATGCA GCTTTTCTC AATGTGTGA CTCTCAGGG  
300 GCTAAAGTC TTTAGCTTA GATTTATCT TTACTAGAGA TCTAGCGAT AGTGGGTTA  
360 ATCACTACA TCTGTACT AGTTATATAG CTTCAGACA TGAAGGAGAC ATCAAAAGG  
420 GATGAGACA ACCCGAGGA TATCGAAGA GGGATGATG AACCCCTTC CTCTGCGAG  
480 GAGACAGG CCAACAGG CACAGACTG AAGCACTTA GATTTTAA GAGGAGAAG  
540 GGGAGCTTT GACAGTCT TCCCTTTTC CAGTTTACC GATTTCTCG CTCTTGCAG  
600 CCTTAGGCC AGTACATTT GCGAATGTC AGATTTTCCC CAGATACTA GAGGAGAGG  
660 GACTTTGGG GGTGGGAGG GGTGTGTGT GTTTTAAAG CATAGTTAC CTGTTCGAC  
720 TTTTAAAG TAGGAAAAA AATAGTGGG CAGGTGAGC ATCAGAGTA AATTTGTG  
780 TTTTATTTT GTCTGCTCT TGAATATGT TCAGATTTG TATGATAGC AGTGAATTT  
840 GATCTCTGT TCGATTAAC ACTATATTT TTTCGAAATG TTACTGTCA AAGGCTCTT  
900 CCGTCCCTT CTTTTCCTA TGTACTCT TCACTTTC TTAATCTG AGCGAGGCA



5 TAGGCAATCA AGAGCTAGAG CATGAACACAG GGCCTTTTCC AAGTAGGCTC TGGGTGTCTT  
AAGCCAGCTT GTGCCCTCTG GTTATAGTAG TGTATATAGAG TCCCTGGCAC CTTTCTTTTC  
AAATGAGCTT AACAGACAG ACTGACGAA GTTATCAGAT TCTTCATCA GATCAGCTAG  
GAGTAGGAG CCGAGGAGT GAGGGGCTTC CTGAGTATTT GCACTGTGCT GTAGTAGCTG  
10 AGTTCTTTTC CATGTTACG AATCTTAGC CAGTTACGT TTACTCAGCA AAGCGTAGA  
TCAATTCAG CATGTAGT CTGGTTGGCA GCGATTGCTA ACGAGAGAA CTGCTCATCA  
GCCAAATCT AAGCCTTCC TTTTAGGAG CACACAGAG AAGGACTTGG TCTCTCTTTT  
CTGTACTTGG TGTACTGCG GTTGACCTCA GCACTGCACT CACACTGGCG AGCAAAAGG  
15 GAGCAGTCA TCTCTTTTCT CTGCCACCC CTTGCCCTTT GTTACCAACA CCAATTTTCC  
AGGGGTACA TGAGTTCTG AATTTTAAA AATGTTTTTT GTTTGGTTTT TTCTGGGAC  
TGATATAGT TTTTAGCAT GTCAATACC CTTCAAGACT CCGAGCTTAG TCAATTTTCT  
GTATTTTCT GTTACAGTA TTTTGTGTG TGTGTGTTTT GCAAGCTAT TTGTGCTGTA  
20 TTTATATTTG ATGATGAT TGATCTCTT TTTTCCCTAA GCGATATGAA TTGTTTTTCT  
TGTATTAT TGTGTTGTG AATAGCTGA GCAAACTGCG GCTGACAG GTFATAGTAG  
GCTTCGAAR CGAGAGAGA GCGGTGGAG TGTACTTGT CTTGACAGC TGACTTACT  
GAGTCTGTA GCTTTTCACT CCAATCTTT GCAAGGCTCA AATGCGACA GAACTCTCC  
35 TCTTTCTCCC ACTGCCATG CGAGGAGCG GACATCTCT ACATCGAACA TGTGTTCTT  
CGAGCCCTTC CCAATGCA GTACCTTTGG GCAATGGGG CATTCATGCG  
GATGCTTGTG TAATCGACA CTTAGCCTTC TCTCTCCCT CCGTCTCTC CCGCAATCA  
CTTCTAGGA CAGCGGACT GTTGGCCAG GGTCTGTTTT CTTGCTAAC TCCAGAGAG  
40 CATCCGAGG CTTTGTACA GTTCTTAAT CCGTTCCCTT CTGTTAAGA ATCATATTTT  
ATAGTAGCTT TCGACGTA CAGTATTCAT TGGGTTACT CTATTATAT CAGTAGCTG  
45 GAATTTGAA GTTGGAGTA GTTAGATCTT TAGCTTTTAT TCTTTATTTT TTGTTATTTAC  
TCTCATGTG TATTAATTTT TGATCTGTT GCTGGCTTTT ATAACTCTA ACGAGAGAG  
50 GAGCACTGCC TGAGCTTTTG CAGATGTTAA TGAACACTG TTTTAAATA AAGAGRGAA  
NKKCAAAA AAAAA  
2416

(2) INFORMATION FOR SEQ ID NO: 298:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 545 base pairs

(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 298:

5 GAAITTCGCA CGAGCCATGC TTGGCTCTC CTGTATCTT AAGTCACTT TGTGGCTGT  
10 TTCTGACTCA GCACTACTCT GCAATTTGCG CAAAGATCA CTAATCTCTT CTTCAGAGCG  
CAAAATGTG GAATAGTCTC TGTCAATGCC TCTCTCATG GGTACACCC TCTGCGACCG  
20 TGTATTATCA GTAAACCA GAGAGAGCC TGTGTGAACT GACTCTGCG AACTGCTTGG  
15 ATGTTTTGTT GCAAGAAATGT AATAGGCAAT CAGCTGTGTT GGTGGATCTG GGGCTCTCTG  
ATGTGATAG AAGGTAAA GGCACCATC TCTTTGACT TTGGGAACT CATTCACAAA  
30 GAAGATTTT CCAAGATCT TCTGAAGATT GGTAAAAAT AGCCGTTTC CAGCCGCTG  
40 AATGATCA TTCTAGATG CTGCTTACC AGGACGAG AACTGATTTA CAGAAATGAC  
ATGAAGAAAT TCAATCCAG AATTTGCAAT ACCTCAAAAT NAAATTTCTAC CTATTAAAAA  
25 NAAAA

(2) INFORMATION FOR SEQ ID NO: 299:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1530 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(1) SEQUENCE DESCRIPTION: SEQ ID NO: 299:

40 GGTCTCTGCT GGCATCATAC TTGTCACTG GTAAACAGTT TCCCACTTA CCGCAGATCA  
AGCTGCTTGC CAGGGCTGTC GCGTCTGTG AATTTGCGAG GCAAGCACT TCCAGAGGCG  
45 TGTGTGCTG CCAAGGATGT GTGGGGCCCC GCGCAGGCTG CTGCGCTGCC GTCCAGGTTGG  
TTGGGCGCAG GCGTGAATCT CCACTGTG GAGCTGTGAT TACTGGAAG ATCATGCGCG  
50 CAGATGATGC CAAGTGCGC GGCATGTGTC ACGGGGGAG CATCTCTGAG ATGATCGAGG  
360 AGCCAGGCGC CATGCTACG ACCGGGATTT GCAACAGCCA GAAACGGGAG GCGTGTGTGG  
CGGCGCTGCG TGTGTGAG GGCACGACT TCTGTGTCTC CATGTGCATC GGTGAGGTTGG  
420 GCAATGTGAG CCGCGAGATC ACTTAAGCT CTAAAGACT TGTGAGGCTG CAGGTCAAGG  
480 TGAATTCGCA AAGATCTCTC AAGGTGCCA AAGAGCTGAC CAATAGGCG ACCCTGTGTT  
540 ATGTGCGGCT GTTCTGTGAG AATGTGACA AGGTCTCTCA GGTGCTCTCT GTTGTGTATTT  
60 CCGGCAACA GCAAGAGGAG GAGGGCGGCA AGCGGTATCA AGCCGAGAG CTGGAGGCGCA

720 TGGAGACCA GTGAGGAC GGGGACATG TCGAGCGAGT CCTCAACCA GAGCGGACCA 540  
5 CTGTGAGCTA GAGCGATG ACCGTGATGC ACTGTGTGG GCTTTCAGAC TCGACCTGCG 600  
AGCGCTTTGT GCGAGGAGGT GTGACATGGA AGCTCATGGA TGAAGTTCGCC GGGATCTGGG 660  
CTGCACGCA CTGCAAGACC AACATCTGCA CAGCTTCGCT GAGCGCCATT AATTTCATG 720  
10 ACGAGATCAG AAAAGGCTGC GTGATACCA TCTGCGGACG CATGACCTTC AGAGGCAATA 780  
AGTCCATGGA GATCAGGTG TTGTGTGAGG CCGACCTGCT TGTGACGAGC TCTCAGAGC 840  
15 GCTACCGGCG GCGGAGTGC TTCTTCACT AGTGTGCTT GAGCGAGGGA GCGAGGTGCG 900  
TGGCTGTGCG CCGCTGTGTT CCGGAGAGG AGGACGAGAA GAGCGCTTT GAGGAGGCA 960  
AAGGCGCGGA CTGCGAGATG AAGCGGAGCG GACAGGGGCA GCGGAGGCT GAGCGCTTGA 1020  
1200 CTGCGCTGTC CTGCGACTG TGGCTGAGT AGCATGCGA AGCGGCGGAG TGTCAAGTCA 1260  
CTTAGAGATT CCGCCGTGCG CCGAAAGCTT AATTCAGATT GAGAGCTGCT GTTGTCTGAA 1320  
25 GTTTTCGAT CAGCATGTTA ACTGTACTC TCTCTGCGA ACTTACACAC CAAAGCTTTA 1380  
TTTATATAT TCGATATCA ATGCTACACA GTTGTGCGG GAGCGCGCGG AGGCTGTGG 1440  
CAGAAAGCTT CCGAATGCT TCGGAGCGG CTGTAGCGTA TCGGAGGAG CAGACAGAC 1500  
30 TATTAAGCT GATCTGCG TCGGAGGAA 1530

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 997 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 300:

(2) INFORMATION FOR SEQ ID NO: 300:

45 AGGTATGTAG AGACATNTA CAGCTACCA ACAGAGGAA GGTGCTGCG CCTTATATT 60  
TAACTATGTT TACAGGGAAT GGTATATG TGGCTTCCG AGTATTCCT CCACATGTT 120  
TTGAGGACC CTAAATCCA AAGCTGATT GATGTGTAA AGACACTTA TAGCCGAGT 180  
50 GAAACCTCT GGGCAGCTT TAGGCTGCA GGTGTGATG CTGGCTCTGT TCCCAAGAC 240  
CCGAGTACG AACTCTTCAG AATGACTTC TATGTCCAG CTGGTCACT GCGAGGCTCA 300  
75 TGAAGGAGAC ATCGATAGG GTGCTCTT TGTCTGCTG TCTGGAAATCC ACCAGCGGCG 360  
TATCTGCTTT TATGGGCTG GGGACTGAA TTGGATGCTT CAAAGCCATC ACCTGTTGCG 420  
60 CAAACATTT GACCGAAGG TAGATATTA TGTCTTTCAG TGTCTTTCAG AATACCTAG 480

5 TTATTAAGGCC ATCTATGGGA CTGAACCTTG AGACACACTA TGAAGACGTT GCTACCTGTC 540  
GGGCAAGAGC ATGTACAAAC ATGCTCAGAA CTGTCTGGGA CAGTGTGGGT GGGAGACGAG 600  
GGCTTTGGAA TTGTGTGCAT CCTTTAGGAT AAGAGGCTG TATATAGATT GTGGTAAAT 660  
AGATCTTTTG CCTTCGAAT TGGTGGCTGG GTGATGCTG CTGTCTCTCT CAGCCCTAAC 720  
10 CCTAGTATTT CCTCCACTAA CTTTCTCACT AAGTAGAAT GAGAACTGCT GTATATGGGA 780  
GAGTAGAGGA GCGATATGTT GTAGAGCACT TGAATTCAGT TGAATGCTG CTGGTATGTT 840  
TTCATATCTG TGAAGCTGCC GTTCTCATTA ATTCAGGTT TGGTACCTG GAGGAGAACT 900  
15 TTGATGAGAA GAGAACTTC CTTCTGTATC TGTTACTTAA AATATATATA GCTCTGATTT 960  
CAAGTAGAG AAAAABAAA AAGAAAAA AACTGCA 997

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2145 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(2) INFORMATION FOR SEQ ID NO: 301:

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 301:

35 TTGAGCGGGA CCGTAGGGCG CCGAAGGAA ACTGGAGGCG GAGGTGACC GGGAGCGGAG 60  
CATTTCAAT CTGTCTGTTA GACTGTGTGC ACCACACCA TTTTGGCTGC AAGGCTGTGT 120  
TGTCTCGGA CACTACCTTC TAGGTTTTC CACCGACTT TCAAGAGCG CTGCGCTGTT 180  
GTGAGAAAT CCAATCGAA GATCAATGG CTGTTAACAC CTAGCAGGGA ATATGCCACC 240  
40 AAAACAGGAA TTGGGATCGG GCGTGGGAGA ACTGGCGGAG AACTCAAGGA GCGAGATG 300  
GAGCATGGA TGGAAAAAT ATTAAAAAT GATCAGATGG GAGATGCTT TGTTCCTGGA 360  
45 GGGGCTGCTG TTGTCTTTGG AGCATGTGCG TACTATGCT TGGGACTGTC TATATAGATT 420  
GGAGCTATG AAGAGCTGT AATTGGCTT CAGTATGTCA AGATAGAA TCAATTCACC 480  
TATATATCT TACAGGGAG TATGTGTTTA ACAGCTTCT CTGCCATAGC AATCAGCAGA 540  
50 AGCGCTGTC TCAATGACTT CATGATGAGA GCGCTTTGGG TGAATATGG TGTGACCTTT 600  
GAGCGCTGG TTGAGCTGG AATGCTGTGA GATCATATAC CATATGACCA GAGCGCGGCG 660  
75 CCAAGCATC TTGCTGTTT TCACTATCT GGTGTGATGG GTGAGTGTGT GCTCTCTCTG 720  
AAGATATAG GGGTCTCTCT TCTCATGAGA GCTGCTATGT ACAAGCTGCG CATTTGTGGA 780  
840 GCGCTGTGCA CTGTGGCCAT GTTGGGCCC AGTGAAATGT TTCTGAACAT GGTGTGACCC 840  
60 CTGGAGTGG GCGTGGGCTCT GGTCTTTGTT TGTCTATGCG GATCTATGTT TCTTCCACT 900

5 ACACACCTGG CTGGTCCAC TCTTTACTCA GTGCAATGT AGGTGGATTT AGTCTTTTC 960  
AGCATGTTC TTCTGTATGA TACCGAGAA GTATACAGC GTGCGAGGT ATCCACATG 1020  
TATGGAGTC AAAATATGA TCCATTAC TCGATCTGA GTATCTACAT GGTATCATTA 1080  
AATATATTA TGGAGTTTC AACTATGTC GCMCTGGAG GCMACGAAA GAATGAGGT 1140  
10 GACTCAGCT CTGGCTTCT TCTTACATCA AATATCTTGT TTATGCGGC AGATATGCT 1200  
TAATATATTT GTACAGGAG CTTCCTTGA AGTTTAGAG ATAGAGAAA TGTCTATTA 1260  
TTAAATATTT CCGTATATGT GATGCTCTAG GTCTGCTTT TTTCTGGAG AATATATCA 1320  
15 GTATCTCTT CCGAATAG CACACATAT TTCAATCTC ATGTTTATGT GATTTTAAA 1380  
TGTTTTGGT AATGTGAAA CTAAATTTG TGTCTATGA ATGTATGCT TTTTCTACT 1440  
20 TTAAATATTA GTAGTTTCA TCGATACTA AATTTAGCA AACCTGTGT TGTATATTT 1500  
TTTGTGCTC AGATATATGT AATTAATGTC ATATGCTAT TGGAGCTTTG GTAAAGGAC 1560  
25 CAGAGAGAG GAGTCACTG CAGTCTTTG TTTTATTTAA TACTTTAGAC TTAGCACTG 1620  
TGTATATAT TATGTAGAG CAGTATAGA AATCTGGGT ATTTGGAAAC AGTGTGCTAT 1680  
TGTATATTC ATCTGCTGA CTTAAGAAA CTGTTTATCC TGAACAGGC ACAGGTATG 1740  
30 CATCTCTG CTGTCTCTC TCGATGCTCT CTTCCTATA TATATGCTGT CATGTTTAC 1800  
TTGTACAGAA TGTATATCAT ACAGAGATC CTGTATGAA TTATATATGT GTGTTTACT 1860  
TTTGATATGT ACGAAGGAA ATACTTTTA ACTATCTC AAGAGAAAT ATTCAGACA 1920  
35 TGAATATAT TGTCTTTTC AGAATACAA CAGTATATC ATGATTTCTA AGTGTTTT 1980  
TATTTTCTCA TATTTATGA ACTGCTAT TGAATACAG TTGCTCTTGT CAGCTCTCA 2040  
40 AGCTTTGAG CTTTATAGA AAGCTCTT TGTGCTTAC ACTGGAAAT ATGAAAGAG 2100  
TTTCTCTCT AAGACTTTG GTTCTGCEA TTGCTCTCA GACTATGAC TAAAGAGAA 2160  
45 AGCAACAG AACTATGCT GTCTTATGA AATATATCA CCGAAGAG TATATAGAA 2220  
ATGCTTCTAT TATTTTCCC TACGAGCTT TTACTTCTCT TACATCTCA CACCTACT 2280  
TCTGTAGAC ATTATATAT CTTTGTATC AGAAGATTA TATTTAGAG GATTTATG 2340  
50 AAGGG 2345

(1) INFORMATION FOR SEQ ID NO: 302:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2369 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 302:

5 TTTTATTTT TTTTATTTT TTTTNCAG ATCATGTTT ATTTATTTT TCGATATAA 60  
AGATATATA CATATTAAGG AATCTCTAA AATTCACATC TAGAGTTATA CACATCTAG 120  
TACTTTTCCA ATGATATGTA ACGACACAA AAAAATCTC TAAACACCTG AAGCCCCAC 180  
10 TTTTACATG GACTATGTA ATTAATAAT TTGCATTTA ATTTTCTCA CATATATAT 240  
TTATATATG AATCAATGG TGTATATCA ATAAATGAT AAGAAATAG TAAATATAA 300  
15 CTTTAAAGG CAAAGTTTA TAGTCTGCA ATCTAATTA TCTTAATTT ATATATAAA 360  
TTAAATATA GAGTTTCTG TTACAAATT CTAAATCTC TGGTTTATA TCAATTTCTG 420  
20 CTACCAATTT AATCTGACA TCTCTAGCA CTGCAATTT ATTTTCTTC CCAATATGT 480  
GTGATATAT AATGATNT TCAAGACGA TATTAATTTA CTGTCTGCA GAAATAGAA 540  
CTCAATATG GTACTGTCA CAACTCTCT CCAATAGAA ATTAATAAA CAGTATATA 600  
25 AATTTTAAA AATCTTAA AAGGATCT ATAGCCAG AGTACATTA ATGTCACAC 660  
TCAAAATC CAGAGACAA TCAACACAC ATCTTTCTC TCTCTCTAG CAGACAGAG 720  
30 TCAATTTTC CATCAATTA CCGATTTGA AGCAGCCAG GGCACACAG TGTACACTG 780  
840 ATTAGATCTT GCAAAATCT AAGATGGAG CAGGCTGCC CAGAGAGAG GTTAAATAT  
ATATATTTA ACTATATAC ACGATATAG GATTCAGCC CATGCCAAC TGCTCTGTG 900  
35 AATCAATGG ACCTTTATG TTAAATAT ATCAATGTA ATATATCAT AATTTATAT 960  
GCGAAGCAA ATCCAGGCA CATTTATAT TATATATTA CTGTCTGTT TCAATTTAA 1020  
40 AATATTTTG CTATATATC ATCTCACTG AATCTATAT AAAAAATCT CTATATATA 1080  
CAGATATTA CTTTGTCTA GTTCAGGCC TTTTGTGAC TTCTGTCTCA ACTATAGCA 1140  
1200 GATCTTACA TGTATATCA CATATGAGA AACTCAAGT GAGTCTATC AATAGCTGTG  
45 CGTATAGCA GCTGTAGGT ACTTTGAAG TCAATATGA CCGAAGCC AATACATTA 1260  
ACATGAGGA TGGCAGAG GATGGAATG CCAATATGC AATTAATCT TTTTATAAA 1320  
CAGAAAGAG AAGCTCTC GTACAGCAG AATCTGTAC ACATATATA ACGAAGAGC 1380  
50 CACCAACAT TTTTAAAC AGAAGCAAT TATATGTTG GAATATATA ATTACAGAA 1440  
ACCAAGATC ACGAGAAA AATCTACTG TTTTCTTGG AGAAGAGA ATGTTTCAAC 1500  
55 CCGAGAGAG TTACTGTG GAGCCGCA CAGACCTCA TGTGTGTG 1560  
CCTTCAGCA TTCACTTCA GGTCTATAT GAGAAATG CCGACTCTC TGTGTAGCA 1620  
60 AATCATCTC CTGTCTGAG AAGAGACA AAGATTTCT CAGGACGTTC GAGCATGAG 1680



5 CCGATGCTGAG CAGTGTGTGCT GTCTGGGACC GTGGGGCCCA CAGAGGGGCT GCTCTCTGCT 960  
GGCACCCTGG CTGCGCTACA CAGTGTCTTC CTGCTCTATC TCGATTCTTC CTACACAAA 1020  
GTGTATGAGG GATCTGTGA CAGCTGTGAG GGGCCGACA TCCCGCCGAT CCGAGGGGTC 1080  
CCGAGAGCA TCGCTGCAT GCTCCCTGCT GCTGGGCTTC CCGACAGGCT CCTGACGCC 1140  
ACAGCCAAAG CTGTGGGCT GACCTGTGAG TCGACTGTAC CCGACTGAA ATCTGTGGCC 1200  
AGTCTCTCTT CCGCGAGCT CAGAGGAG GAGACTATTT AAGGAGCT CCGTATGACA 1260  
TGTTCTGAG ATGGGTTTG CAGCTGTGAC TGGCTGTGAG CTGGGTAGT ACCTCTTGN 1320  
AGCTGTGGC ACTTCTGAA GCGAGGGC AGAGCTCTT GCGAGGACT GCGAGGCTCT 1380  
GCGAGCATG CAGAAATGG GTGAGCTCT TCGAGAGCC GTCCGACCT ACCGCTCTCT 1440  
TCTCTTTAT CTCTCCACA TTCTCTTCT AATATAGAG TTGCTAATTA AAAAAAAAAA 1500  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA GGGGCC 1537

25 (2) INFORMATION FOR SEQ ID NO: 305:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1493 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

30 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 305:

35 TCGATCCCA AACGATGCC TCGCAACAA ATCTTAGAC ATCCCATAT ATATATGTT 60  
TTATATTTCT ATTACATCA TTATTGAAA TACCGAGTC AGTCCCTGC TTATATAGT 120  
TTTATTTCC TTACTTACT TTCTCTATTT TTTTATTTG AATGGAGAT GAGCAAAAT 180  
ACACATTCAT GGTGAGCA ATTTTGTGA CATTCTTCT TACCAAGA TCTATATCA 240  
GGATGATCT GAGCTGTCA ACGAGCTGT ATATAGCAG ACATGAAAC TCTTTGCGA 300  
GCTGGAAAT AAAGGAAG AGAGAGAG CAAAGAGT CATGAAAG AACGACAAAG 360  
GGAGGAGAG ATTGAGCTC AGAGAAAGC CAAAGCGAA AGAGGTGCG AGAAATCTT 420  
TGAGAAAGT CAGATGTC CTGTGAGAG CTGGGAAAC TTCCAGCCA ATACAGAGG 480  
GAGAAAGAG AAAAAAAAAA GAGCTTCTT GAGCCAGC AAGTAAAAA TGGAGCAGG 540  
TGAATGAGC CCAAGTCA CAGGACAGA AGCTTTGCC TCGTATCTC CTCTCTCTT 600  
CAGAGGCTC ATTCTTCTT CCGACTTCA CCGCACTA CCGTATATT TCGTTTATG 660  
TCCATTTCT TTCAATAG ATTATATC GATCAGATTA ATCTTTTCT ACATGAAT 720  
GAGGGGCTG GTTAAAAA AGAGCTTTCC CTCTCCCTGC CCGTAGACA ACCAGTTTA 780

5 GAGGTGCA CAGTGTGTC TCGTCTCTT TCCACAGCC TGTAACTCAG TGTTTTATC 840  
TTCACTGAT TGTATGTTT AGAACTCTG TGTATGTTT GTGGAAATCA TGCANTAAA 900  
CTATCTCTT AAAGCATGT TCTGTGACT TCGAGAGCA GCTTGGAAAG GCGACTTAG 960  
GAGGCGGCTT CCGTTCACTT GCTGCTCTCT GGTGTGCTC CTTTGGAGG CCGAGTATG 1020  
10 ACAGGAGCA CTGTGAGCA CAGAGCAG CAGTGTATC CTTGTGTGT TTAGCATGTG 1080  
CCGCTCTCT ACTGACCAT CAGTGTGCA TGAAGGCCAC GCGACCCAAA CTTTCACTT 1140  
TCCAAAGAG TAGCGTCTT CCGCCCATTA CAGTGTCTA GCTGTCTGC ATTTGTATG 1200  
GTTATATTC TTATGTATA ATAAATTTT ATACCGAGC CATTGATGA CTTTCTCTG 1260  
TACTCTCTT TGTGGTCCC TTGTCTGCT TGGCTGAGC CCAAAATCT TTGGGTTGG 1320  
20 ACAGGCTGG CTGAACCTTA GTTCTCTCAT CTATGAAATG GGAATATGAA TTAAGTCCAG 1380  
AGCTTTAGG GCGATTTGC CATGCTAT ATAGGTAAAC TACATAGTG CTCTGTGCT 1440  
ATTGCCATA TCTATATTT TCTGTGAAA ATGAGATAC TGAATTTTT GAG 1493

30 (2) INFORMATION FOR SEQ ID NO: 306:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 577 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 306:

40 AATTGCGAG AGGATATTA TACACTATAC TGGCATTTAC TGTTCACCC AGCCCGGAA 60  
GTCAGAGTG TATATGGAA AATTAGAAC TCGATCTACA TTGTTTCCA GAGGCTCTC 120  
ATAGCATTT ACCGAGAT CTACAGCAT GATTAAGACA CTTATATTC TTATCAACT 180  
45 GACTATATC TATATTTTA TTTTATTTT TGTGTATAT GCGAGCTAC TTCACTCTT 240  
AAACTCTT TCAATTTGT ATGTAAACT TTAAACATG CAGATCAGT TAGACTGCT 300  
CATAGAGAA GAGCTAGAA TCGAGTACA TGAATTTTA ATACCTCTC TTGTTTTTG 360  
50 ATGTAAACA GTAAATCCA GTACTGACA AGAACAGT GATTATATC ACTATCTG 420  
AGGATTTCA TTTTATATC ATCTTTATG AATTTAGAA CTCATCTCT GTTTTAAAG 480  
55 GGAATTTTA ATTAGAAT AACATTTGT GACAAATG TTAATAAAA AAAAAAAAAA 540  
AAAAAAAAA AAAAAAAAAA AAAAAAAAAA AACTGCA 577

60

## (2) INFORMATION FOR SEQ ID NO: 307:

## (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2660 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 307:

5 GTTGTGACCG CTCCTGCAAT ATGGCTCCCG CGGGCTGCGA GRRKCTGRT CNCRGTGGC 60  
120 TACGCTGTTC TGAAGAGGGA GAUTTAAGCT CCGTTCCTCC ACCGTGCGCG CTGGCCAGGT  
15 GGGCTGAGGG TGAACGAGAG ACGAGAGCT GCTTGTCTGA GCTTAATGCT CAGAGCTGG 180  
240 GAGGAGGTT CCGCGGCTTC TGTGCTGTCA GCGCGGCGAG CCGCTTCGCG CTTCATTTCC  
20 TCCCGCAGCC CCGCTACTG AGAAGCTCG GANTCCGAG AGCTGGCGAG CCGTGGCTTC 300  
360 AGCTGCGGG GCTTCAGTC AGCGGAGAG CAGCGGCGAG TGGGAGGGA GAGAGAGCC  
420 TTGACATCTC CAGCTGACCA GAGGTCTGCG CTGGAACCGA GCGAGCTGCT CCGTCTAGGA  
25 TGACTGAGC CTGAGCTCT CAGTTTTC ACGTGGAGC ATTGAGTGA GCGCGAGAG 480  
540 ATGGCTCTGA GCGGAGCAGA GGAAGCTGCG ATTATGGAG CCGGCTGCTT CCGATGAGT  
600 CAGAGTTTCA GCGGAGAGC GCGAATTTG CCGCTTGA TAAAGTTCGA CCGTCAACTA  
30 CCGAAGGGA AGAGTGTCCA GTACGCGGA TCGAAGCGA TTGAGCGGAG ATCGGCTTT  
660 CAGTGGTTC TCGCGGGTTC TCGCGGAG TGTGCTGGA CTTTCAGAGT ACTGAGCA  
720 GAGCAGAG TACTGACCG ACTTGGAA TACAGAGAG GCTTCAGAG GTACAGCGCC  
35 TATGAGGCG TGTCTGAAA CTTTAAGAG GGGTCAATG CTTGCAATCT GCGACTGTG 840  
900 CAGATCGACC GGGACTCTCG CAGTCTGAG CCGCTTGTAA ATGCGCAGTG CAGAGTGA  
960 TATTAAGAG GCGAGAGCG TGTGAGTTC CCGATTGGA AGAGAGATC TCGAGTGT  
1020 GAGCTCTCTG GTGAGAGTG GCGCAATGT GATGCGCGG GTCTGCGGG ACTTCTTCA  
45 GAGGCGGGA GCGACTGCT TTATTTGCG TGAAGTACCG CTCTCTTTGG CCGCTTGCAC  
1080 CAGAGTGG GATGTGTAA GCTTACTCTT GAGAGACCA CACAGCGCG CCGAGCTGCA  
1140 GCGCTGAG TCGCAGGGA AGAGTGTCT GATGCGCTA GTGAGTATC TCGAGCACT  
1200 CAGCTGAG CAGTGCATG GTGACAGCA TGTATGATG GCTTCTCCAA GCTTGGGCG  
1260 GCGTCTGCG CTACGCTGCA GCTTGAAGC ATCGGAGCC TCGAGAGTCT CAGCGCTG  
55 AGCTGCGCG CCGAGGAGG CAGATGAG ATTTCAGCG ACATCTGCA GCGGAGTTT  
1380 TGAAGTGA GCGACTTTC CCGAAGTTC ACCGAGTGT GCTATGCGCC TGTGCGGTG 1440  
1500 TCGCTGATG AGCTGGCTTC TGTGAGAGC ACTGAGTGT GAGATGATT

1560 GCGTTTCAAT GCAAGAGCC GCAAGAGAG CCAATGCTCG TTTTGGAGCC CCGAAGCAA  
1620 CTCTGCGAG CCAATGGA TCTGCTGAT CCGAGTTCT TCTTAAGCTT CTTGTGTAT  
1680 CTGATCTACA TTTTCAATTT CAGCGCTGTT GCTTACCATC AGCTTACCTT GAGAGAGAG  
1740 GCGCGCCCTC ACTGAAAGC GAGGTTTGGA AACTGCAATC TGTGAGCGG CAGATCTTT  
1800 ATCTGCTGAG GGGGATCTA CTTCTCTGT GGGCAGCTG TGTACTTTCT GCGCGCGCA  
1860 GGTGTTCATC TGAATCTGTT TCAAGAGAG CTACTTTTGA AATCTCTTTC CTTTTCAGG  
1920 CCGCTCTTCA CAGTGTGTTC CCAAGTCTG TTTTTCCTCG GCAATGAGT GGTACTGCG  
1980 CTTGCTTTTG TCTGCGCTCG TGGCTGGCT GGTGAACTT GCTTTACTTA TACAGTGGC  
2040 GTTCCAGCAC AAGGAGATC TACAGTTTCA TGTTCCTGA AGCGCTGCTG AGCTGAGCC  
2100 AGAGGCTGCG GCGCGCGGA GCTCTTACAG GCGCGAATTC CAGAGTGA GTGAGCGCA  
2160 TGAAGGAGCA GAGAGAGAG GCGAAGCGG CCAATACAG GGTATCTG GAGGCTCTCT  
2220 TCGAGCTTCT CAAATTCAC ATCGCGATG GCGAGCTGCG CTTTCAGAG CAGCTGAGT  
2280 TCGCGGCAAT GGTGCTCTCG CTGCTGCTG CTTACCTGCT GCTACCTTAC ATCTGCTGCG  
2340 TCAAGTGTCT CATGCGCTTC ATGAGCGGA CPTGAGTTC TCGCACTGA CAGCTGAGC  
2400 ATCTGAGAG TCGAAGAGC CAGTCTCTTC CTGAGATG AGAATGCTA TTTGTGCTGCG  
2460 AGCAAGAGC AGCGCGGAG TGTATGCTG ACCTTTGCA CTAAAGCGAG ATGCGAGCC  
2520 CAGTAGCGC TGGTCTTCA GGTGAGAGA GGTCAACTG GCTTCAATG GAGCAGAGC  
35 TCGCTTACCT GTGTAGGAC CCGTGAAGG CAGGTGTGCG TCGACTCTC GAGAGCGCTG  
2580 TCTTGGCTTC CCGTCCGAG GAGATGAGG ATGGTGGCTC TGAAGAAAC TATGTGCGCG  
2640 TCGAGCTGCT CAGTTCGAC TGTGCGCCA GATCGAGCG GAGCGCGAG CAGCAGAGAG  
2700 AGGATCTTTC CAGCCATC TGTGCTGCT GGGTTCGCG TGAATTTGCG TGGCAATAT  
2760 ATATTTTCAG TAACTGAAA AAAAAAAAA AAAAAAAAA AAAAAAAAA GCGCGCGCT  
45 ASCCAATTC GCGCTTATG TGAATGCGA TTACGATAA 2860

## (2) INFORMATION FOR SEQ ID NO: 308:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 876 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (1) SEQUENCE DESCRIPTION: SEQ ID NO: 308:

CTGCTGTGTCT CTGCGCTGCTCT CTGAGCTCTGCT TTCTACTATPAC AGCTGGCTCTC 60  
CAGCAGCAGG GCATCTACAG TGTCTATCATC CAGAGAGCTCT GCTGAGCTCTG AGCAGAGANN 120  
5 TTGGCGCCGCC GAGAGCTCTCTA CAGGCGCCCAA TGGCAGCAGAG TCACTGACAG CCAATGAGAGG 180  
ACAGAGAGAC GAGGCGAAGG GGGGCCAGTA CAGGGGTATC CTGGAGGCTT CTTTGAGCT 240  
CTTCAMATTC ACATCTGCTA TGGGAGAGCT GGGCTTCCAG GAGCAGCTGC ACTTTCGCGG 300  
CATGCTGCTG CTGCTGCTGCT TGGCTTACCT GCTGCTGCTGCT TACATCTGCT TCTTCAGAT 360  
GCTCATCTGCC CTCTATGAGC GAGACGCGCA ACHATGCTGCT CACTGACAGC TGGAGCATCT 420  
15 GGAAGCTGCA GAAAGCCATC TCTGTCTCTG AGATGAGGAA TGGCTATCTG TGGTGCAGAA 480  
AGAGGAGAG GGGCAGGTGCT ATGCTGACCG TTGGCACTAA GGCAGATGCG AGCCGCGATG 540  
AGGCTGCTGCT CTTCAGGCTG GAGAGGTGCA ACTGCGCTTC ATGGAGGCG AGCTGCTGCTA 600  
GGCTGTGTGA GAGACGCTCA GGGGAGGTG TCCCTCTGAACT TCTGAGAGAC CTTCTCTCTG 660  
CTTCCGCTCC CAGAGGAGAT GAGATGCTG CTCTGTGAGGA AACATATCTG CCGCTGACG 720  
TCTCTGAGTC CACTATATG CCGAGATGCA GCGAGAGGCC AGAGGAGAGA CAGAGAGATC 780  
TTTTCAGCA CACTGCTGCT CTCTGCTGCT CCACTGATTT CTGCTGCGCA ATATATATTT 840  
30 TCACTAATMM AAAAAAAAAA AAAAAAAAAA ACTGGA 876

## (2) INFORMATION FOR SEQ ID NO: 309:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2025 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double  
(D) TOPOLOGY: linear

## (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 309:

CATGACCCGC CTGATGCGAT CCGGACAGC CTCTGCTTCC AGCTCATCTT CTCTGATGAG 60  
45 CAGCCGZAGC GCTTCCGCA CAGGAGAGG CAGCCGAGGC GCTTCCGCA CAGGAGAGG 120  
CAGCCGAGC GCTTCCGCA CAGGAGAGG GCGCCAGCTG GACATCAGC CCACTGCGG 180  
50 TGTCTGCTGG AACAGCCGCG GCGCAAGTCC ATGAGAGTCT CCTGCTAGGC GCTTCCGCA 240  
GCTGCGCCGC CCGAGCTCTG ATCTCTGTAG TGGCGCCGCT CTCCCGCGCC CTTTTCGCGC 300  
CCCTGCTGCT CATTACTGCT CTAAGCTGCT ATTAATCCAA AGCTTATTTT GTAGAGGTGA 360  
GCTCTGCTGG AGAATATGA GCTCTATTAAC GTGGGTGCCC TTCTCAAGG CCGGCTGCGG 420  
GTGACACAAA GGAAGGAGC AAGCATCTCC GCATGCGATC CTCTTCATTT AACCATGCGC 480  
60 GGGTTCGAC TCTCTCTCCC TCTCTGAGG AACCAAACTT GCGAAAAACA AGACGGTAC 540

AGCACACACT TCAAGAGGC AGGCTAGGC CCGCTGAGC ATCTGTGTTT AACGGGTGCG 600  
5 CTGGTCAGAA GGGCAGCGCC CCACTTCCCG TTCTCTCTTT AACGAGGAG AGCTGATCC 660  
AGTTTCGCGA AACAAATCC TTCTCTCATT TGGGAGGGG GGTATATGCT ACATGAGGCG 720  
ACCTCTTTTA AACAGCGAA ACAGAGAGG GGAAGAGGTG GATATCATCT CAGGCTAGA 780  
10 GGCATTTGGA ACACATATC TACGTAGTTA ACTTGAAGAA AGCATTTTTT AAGTTGTGTC 840  
CATCTAGAAA GCTTTGAATG CAGAGCGAAA CAGCTTGAT TTCTCTAGCA TCGCTTAAAT 900  
GTGAGCGAAA AGCAGGCTAC AAAATCTCTT GCGTTTACAG ACGAAATATTT TTGCGAAGC 960  
15 GTTGGGCATC ATGGTTTTTG AAGGCTTTAG TTCTGCTTTC TCGCTCTCTT CACAGAGCCC 1020  
AAGCTCTCAC CCGTATATCA TGAAGCATG ATTAATCTTG TTCAAGGAGA AGATCATTTA 1080  
20 GATTTGTGTT GATTTGCTTA GAATGAGGG CACATATCCA CAGCTGCGCT GCGTGTGATG 1140  
AGTGTCTCTG CAGGCGCGCG AGTAGAGCA CTGGGCTGGG GCGGAGATG GGGTTACTCG 1200  
ATGTAAAGGA TTCTCTGTTG TTGTGTTCAG ATCAATGCA GTTGTGATTT CTGTGATCC 1260  
25 CAGCTTGTGT CAGGATTTT TGTGTATG GGTAAATCC AGTTTTCAT CTTCAGACAG 1320  
TGGGCTGAAA CGTGACTCA GTAGCTGAA CTTGTGACC GGTCACTTT CTTCGATCTT 1380  
30 CAGACTCTT TCGCTTGTTC GGGTGGGGG TGGGAATCCA CTTGGGAGC GGTGCTGAG 1440  
AAATGTATAG GATTTCTGAA TACATATTC ATGGGACTTT CTTGCTCTCT CTTGCTCTCT 1500  
35 CTTTTCCTGC TCGCTTACTT TTCCCGAT TTGGGAGC CACTGATGTT TCTGGCGGCG 1560  
CAGTCCGCTT GCGAGTTTC TGTACTAGT CTTTGTACTT TTCAATTTTG CTGACCTGCG 1620  
ATTTTCTCAT AGGAGTTTG GTGAGATGCA ATGATATAT GTATGCTAGC CACTGGAGC 1680  
40 CAGGATTTT TGGAGCCCG CAGTTGGGAG GAGGAATGAG TCCAGGCTTC CAGGTGCGCT 1740  
GAGAGCAT GACTGTTAC CTGCGGCCA TCACTTGA GGGTTTCTCT GCGCTTGAAT 1800  
45 AAAAAATGCG GGAATCGGG CAGAGAGGC TGAATAGGGA TGGGAATTA TTGTGACAA 1860  
GTCTTTCAG AGAGTTTCT TATGAGATA TTGTGATTTA TTTCAGAGC ATTAATATTTG 1920  
TAATTTGCA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA 1980  
50 GAGGCGGCGC CTTACCCAT TCGCTTATA TGAATGATAA CATC 2025

## (2) INFORMATION FOR SEQ ID NO: 310:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3026 base pairs  
(B) TYPE: nucleic acid  
(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 310:

5 TAGGACAGAC TGAATATCC TAAACCCCTTA AGCTCCAGGT GCCCTGTGGA AGCAGCAACT 60  
GACTATAGC AGGCTGAGGC TGTGTCTTCC TGGTCATAGG CTCACCTCTTT CCCCCAAATC 120  
TTCTCTGGA GCTTTTGAGC CAAAGTGTCTA AAGAGAAATAG GTAGAGAAC TCTTCTATCT 180  
AATCTTAAA ACCGATATGT TGAATCTCA TTCAACAGCT GATGCCCTAT AACCCCTGEC 240  
TGAATTTCTT CTTATATAGC TATATAGAT AGCAAGATCT TTACATTAAT CAGAGTGGTT 300  
TCAATGGCTT CTTACCTCTT CTAATGGCC CTCCTTTTAT TTGACTATAG CATTACAG 360  
TGGACTAGC AATTATCAA GATATATAGA AATACAGTGC TTATATGGTC TAACTATACT 420  
GCCCTCATA TCAAGCTGC CTGGGAAG GATGCGAGCC TCAGGCTTC CTTATGTGCT 480  
CCACACAG AGCTCTTGA TGAAGTCAAT CTTTGTCCC TATCTGTATC TTCCCTGCC 540  
GCCCTCAAT GGTATATGG TACCGAGCT GGTCTTTGGG CTAGATATG GGGACAGT 600  
TCAATACCT CTTATAGT CTATAGAT AACTATGCT AACTATGTT ATAGATATTA GTGGAGAG 660  
CTGGCTTTC CTATATAG CACTGATCC TACTCTTAC TGGTCAGCC GCTGCTTCA 720  
GGTATGGAC CTCTAGT TGGAAATAC TGAATAGGA GAGGAAATA CAGAGAGGEC 780  
CTCTGTGTT CTTGCTCA GCGAGTGG CACAGGCAAT AACCAATTA AACAGATTA 840  
CTGAGTCAAT TTTTATAT GGTCTCTTC ATTCGACTG CACTTGTGTC TGGTTTGCT 900  
GACTGGGAC ACCGATAC TACAGATCT GACAGGAGA CTGGAGCTG TCGACTCTA 960  
GCTGGAACT TACTGTGTA AATAATCTT AGAATCTCTA CCAATAGTG AATATGCCAC 1020  
ATTTGCTTT AATATCTTA CCAATGTGG GAATACTGG CTTTGTCCA GCTTCTTCA 1080  
GGGCAATAAA CTCAGCCCT TGGATAGCA GTCCGATAG CCAATATTT TTTTATAGAA 1140  
AACTTCACT TGTTTTCT TTACAGTGA CTTCTTCTT GCCCAAAAT TATATACTCT 1200  
45 AAGGTATAAA AAGATCTTA ACACAGCTT CTTGCTGTA AATATATTA TTATATCT 1260  
GTATTTTAA ATCTGCTCC TGAATATCA CTGTCCCTAT CTCACCTAC TGCATTTGG 1320  
GCCTTTCCA TTGCTGCA TGTCTTAT CATTGAGC CAGTGGAG AGGAGAGG 1380  
GAGAGAGGG GTCCGACA CTTGTGTTC TTCTGACT ATCTGTACA AGAAGATTA 1440  
ACACTAGGC GCTGCTGCC ATGCACAT CTCGAAACA CTTATCTCC TCGAGAGTG 1500  
55 GCTTTGAG GCTCTTACT GGGAGCAT TAAAGCCCT CTTACCCCT TCTTTTTC 1560  
TTCTTACT CTTTGGCT CAAAGATTT TGGAAAGAA ACAATATCT TTACATCT 1620  
TTTCAATTC TAAATTTCA GGGATCTG AATATAGG CAGTGGCT AAGCTGCTG 1680

TAAAGTTGAG GGGAGAGAA ATCTTAGAT TACAGATTA AAGAGATC CCTATACAA 1740  
AAGAGCAAT AAGATGCT TTCAATTTG CAGCTTTCC TTTTCATGAC AGCTACTAC 1800  
5 CTGGAGAG TAAATTTTA TTACCAAG AAGTGGTTC ACTGACCTC TGAAGAGCTG 1860  
AATATCAG CACTCTCAT CACTCTACA GATCCAGG AGTCCGAG AGTCCAGCT 1920  
CCTTAACCT AGCTAGTAA ATAAACCTGG CCAAGTGGAG CAGAGCAAT GAGGAGAT 1980  
10 CCACTCTGA GGTATAGGC AGCATGAAA GCAAGAGAG GAAAGATTA TCAAGAGGAG 2040  
AAGAGATC ATTATATGG GTCTGAAAG AAGTCTTT CTATCTCCAC ATCTACTCT 2100  
15 AATACCTGTA AGCTTTTAT GTCCAGAT GCAAAAAA ATCAGTAT GGTATATTA 2160  
TAAATCTCTT TCCCTGGAGT CAGTTTATTT AAAAGATTA CTCTTAGTT TTACTTCTT 2220  
AATTCATAAA CAGAGAGGAG CTAGGCAAT TCCCTGTAGG AATTAAGATA AAGAGATAG 2280  
20 AAGAGATTA AGCTCTAAT AGATCAGAG CTTTCCAGG TATTAACCT AATATTAAGA 2340  
AATACATTA CAGAGAGTGG AATATGAT AGTCTCTAT AGCTACCCAC AGAGCAAGT 2400  
25 ATTATTAAT TTCAATCCA AACTACTTTC TTATATCAC TTTGCTCTCC ATTTTTCOA 2460  
GGAGAGAAA TATGTCCCC CTTACTTTC TTCTTCTAAA AATTAATTC CAGATCTCA 2520  
AGATCATCT ACAGTAAAT TTGCACAGAC ATCTCTCTAC CCAATGCTT GTCTGAGCT 2580  
30 CAGCCAGCT CAGCCAGCA ACTTGTGCT GAGCCCACT GCTTAACT TCTGAGGAG 2640  
GGGATTAG TACACTAGA GACCAGAG TGAATGCGAA AGGTTGAGA CTTCAATG 2700  
35 TTGGCTGTC AGAGTTGCG CACTGCTAT GTTCTTATTC TCGACTGTG AATTTGTGT 2760  
GGAAAACT GTGGTTGTG CTAAATTTCTG TCCAGAAAT AGGTTGACA CAGCTGTGT 2820  
GGTCTGATG AGGATTTGG ACTGCTTAT GTTCTTATTC TCGACTGTG AATTTGTGT 2880  
40 ATGTAAAAA GAAATTTCTG TAAACTTAT GTCTGTATTA ATATGAGCG TTAAACAGT 2940  
AATATTTCA ATAGAGTGC AAAAAAAA AAAAAAACT CAGAGGAGG CCCGTTACC 3000  
45 AATTTTCAA ATAGAGATG TATTAC 3026

(2) INFORMATION FOR SEQ ID NO: 311:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 712 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 311:

60 CAGGCTTGG TGTCTACTTA CAGCTGGGT GAGCAGGGTG CCAAGCACTT GTTCTCTTT



CTCTCTGTC ACCAGGGCT TTCTGCTCC GATTTGGAC TTGTGAMTGG TTGTGGGTGCT 120  
 GTGGTCTGCT CCATGCTGG CTCTGCTCC GTGTGGACT TCGTGGCCAA GCACGTGGAA 180  
 CTGCTGCTTC TTGTGAGTTC GATCTGTGGC TTCCGCTCTGG GGGGCTTACG CTGTGAGACT 240  
 GCGTTGGTCT TCGACCTGGA CACCTTGGGG GCGACGATGG AGCGTGGCAC AATCTTGAGA 300  
 GGGTCAGGCT TCGTGAAGCT ATGTCTGGAG CACTTCTTGG GAGGGCTGGT CACGAGATGC 360  
 ACTTTCATCT GAGTATGG CTGAGCGAG CTGGCCGCCA GGGCTGTGAG GCGACGACT 420  
 ACGAGCTTCT GGGCAGCTGG GAGCTGTCTGG GGAAGCTGCT GCTTGGGACT CTGTGGAGGC 480  
 CTGGCTGATG GGTTGGGGCC ACATCCCTGG TTCTGTGCTCC TCGTCACTCT CTCTGCTTTT 540  
 CCGCTTCTGT ACTTGAAGCT AGCAGCGAG ACCTTCTTCT GAGCTGAGTG GCGTGGAGTGG 600  
 TCATTAAGC CACATGTGCC TGTGGCCGA AAAAAA AAAA AAAAAA AAAAAA 660  
 AACTGGAGGG GGGGGGGGGT ACCGATATGG CCGATATGA TCGTAACGA TC 712

25

(2) INFORMATION FOR SEQ ID NO: 312:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 1289 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: double  
 (D) TOPOLOGY: linear

30

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 312:

CAAAATTTC GACTTTCAG GAGGGCAGA GATATACAA CAAGATTTC TCGAGTATTT 60  
 TTGCAAGCT TGTGTTGAG CTGCAAGAA ATATTATGG TGAAGACTTT TCTGTTTCC 120  
 GTTATGGGT TTTTGGTTGG TTTTGGTTGG TTTTGTACTA TCGTGTGCTC TGTAAATA 180  
 TCGACTGAA CTACATTCAG AAGGAATAT TGTCTACATA GAATATTATA TGAAGTGGT 240  
 ACATATTTCT GATGAGGAA AAAAACTTT GCAATCTTT AGCGATATTT GTTGTTTTC 300  
 TGTGTGTTT TCCCTGGATG AATATATCG TATTATATG ACAGCATATTT ATTCAAGTGT 360  
 TTGAGCTTAT TATTTGTTTC TTGTCTGTA TTTATACATA TGTGTATTTT GGAAGTATTT 420  
 GCGTTTATTA AGCGAAGCTA TATTTGATA CATATGATA AAGCGATGG TGAACCCCTTT 480  
 GTGCTCTTC CACTGAGAT ACACAGAGC ATGTATATCC ATTCTCTTGC ACTTCTTCT 540  
 TCTTATCTTG TTATTAAGCT TTATTAATTT TTGTAGAGGG ACAGGAGTG GCGAAGGGA 600  
 AGAGAGACT TATTTGACTA ACCAGCCCT CTGTGTCCA CAGAGCTTTT GCGTGTGGG 660  
 GAGGCTCTC ATCAGAGGG GCGCAGAGG GGAAGAGAA GTGGGCGAA GCGTGGCTTC 720

60

CGCGCTGGG AGCTTTGCA TCTGAGCAC GCGTCTTCCA GCGCATGCTC CTGGAAGTGG 780  
 GAATGTCAA CCGAGCCCT TACAGAGCC CTCCAGATC TATAGACTT GATCTACTC 840  
 5 TAAAGCAATA TTTAATCCA CCTCTACTCA TTGTAGCTCA GTCCAGAGAC TAAAGCTTGA 900  
 ATGGGGTGT TCGAGCTTTC AGCGAGATGG CCAAGCGGTC CCGTGGGGG TGTGGAGCG 960  
 GCGTATCTCT TCTCTGTGCG CAACTTTGCG GTCCAGACTC CTCCGCGCCC ATCCGCTGAC 1020  
 CCGTCCGCG TCTGTGTCTG TCCATAGCTG TGAATGACG TGAAGAGCA AATAGAGCC 1080  
 CGTGGGCCA GCTCGAGAGG TCGCTGAGCA AGGCTCCGAC GTCTCCGAG TCGAGCCCTT 1140  
 15 GGAATGCAAT TCCGTGTGCT GCGTATTTCC TCGAGATCT GTATAGGCTT GCGCTATAGA 1200  
 AATATAGCT CTTCATGCTG TATTAAGAG ACTTTTAAA GCAAAAAA AAAAAA 1260  
 CTTGAGGGG GAGCCGCTAC CCAATTTC 1289

(2) INFORMATION FOR SEQ ID NO: 313:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 22 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 313:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Ile Leu Phe Ser Ser Ser 15  
 1  
 35 Leu Pro Phe Leu Trp Leu 20

(2) INFORMATION FOR SEQ ID NO: 314:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 128 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

Met Met Phe Leu Thr Gln Gly Pro Leu Pro Ser Thr Arg Ala Arg 15  
 1  
 35 Pro Thr Cys Gln Ala Gly Ala Leu Pro Lys Pro Ser Gly Leu Leu Gly 30  
 45 Val Thr Cys Trp Asn Gly Leu Lys Gly Pro Leu Cys Gly Asn Arg Cys 45  
 55 Ser Pro Asn Thr Leu Leu Ala Ala Arg Gln Ala Leu Trp Lys Gly 60  
 60 Arg Gly Arg Thr His Gln Asp Leu Pro Gly Pro Leu Gln Gly Arg Gln

65 70 75 80  
 Leu Gly Pro Glu Pro Lys His Leu Ala Leu Leu Pro Pro Arg Gly Gln 95  
 5 Glu Ala Ser Trp Ala Ser Ser Leu Pro Gly Gln Gly Pro Leu Pro Leu 110  
 100 105 110  
 Pro His Ile Asn Cys Thr Val Phe Ser Leu Lys Ala Ser Phe Ile Lys 125  
 115 120 125

15

## (2) INFORMATION FOR SEQ ID NO: 315:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

25 Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu 15  
 1 5 10  
 Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser 25

30

## (2) INFORMATION FOR SEQ ID NO: 316:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

40 Met Asp Gly Phe Ser Ser Arg Leu Phe Ser Ser Leu Pro Phe Val Ala 15  
 1 5 10  
 Leu Gln Trp Phe Ile Val Ile Ser His Leu Leu Ser Leu Ser Leu Ser 30  
 20 25  
 Ala Cys Cys Tyr Gln Thr His Cys Ser Leu Xaa Gln Leu Ser Ser Ala 45  
 35 40  
 Phe Ser Xaa Met Gly Glu Ser Cys Val Gly Glu Arg Glu Tyr Xaa Phe 60  
 50 55

55

## (2) INFORMATION FOR SEQ ID NO: 317:

## (1) SEQUENCE CHARACTERISTICS:

60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Ile Ile Thr Pro Phe

(A) LENGTH: 21 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 317:  
 5 Met Pro Leu Ile Asn Leu Leu Leu Tyr Tyr Val Pro Asn Gly Gly 15  
 1 5 10  
 Lys Gln Asp Lys Lys 20

10

## (2) INFORMATION FOR SEQ ID NO: 318:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

20 Met Gly Arg His Leu Val Leu Val Met Phe Ile Thr Thr Ser Leu His 15  
 1 5 10  
 Ser Gly Thr Pro Val Pro Glu Asn Val Ile Cys Gly Val Thr Lys Gly 30  
 20 25  
 Pro Gln Gly Lys Lys Lys 35

30

## (2) INFORMATION FOR SEQ ID NO: 319:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 319:

40 Met Leu Trp Trp Ser Arg Asp Tyr Thr Met Val Phe Leu Leu Phe Thr 15  
 1 5 10  
 Met Val Phe Thr Gly Asp Leu Val Ile Arg Gly Arg Thr Glu Leu Ser 30  
 20 25  
 Leu 30

50

## (2) INFORMATION FOR SEQ ID NO: 320:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 88 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 320:

60 Met Val Cys Ser Ser Leu Cys Asp Ile Gly Ile Ile Thr Pro Phe

1 Ile Val Phe Arg Leu Arg Glu Val Trp Gln Ala Leu Pro Leu Ile Leu 15  
20 25 30  
5 Phe Ala Val Leu Gly Leu Leu Ala Ala Gly Val Thr Leu Leu Leu Pro 45  
35 40 45  
10 Glu Thr Lys Gly Val Ala Leu Pro Glu Thr Met Lys Asp Ala Glu Asn 60  
50 55 60  
Leu Gly Arg Lys Ala Lys Pro Lys Glu Asn Thr Ile Tyr Leu Lys Val 70  
65 75 80  
15 Gln Thr Ser Glu Pro Ser Gly Thr 85

## (2) INFORMATION FOR SEQ ID NO: 321:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 321:

25 Met Gln Pro Gly Ala Gly Val Leu Val Leu Gly Leu Leu Leu Pro Pro 15  
1 5 10  
30 Pro Gln Ser Pro Ser Leu Ser 20

## (2) INFORMATION FOR SEQ ID NO: 322:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 322:

40 Met Thr Phe Thr Leu Gly Asp Ser Gln Val Leu Leu Ile Asn Leu Phe 15  
1 5 10  
45 Pro Ser Met Pro Ser Gly Ser Cys Ala Arg Pro 20  
25

## (2) INFORMATION FOR SEQ ID NO: 323:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 64 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 323:

50 Met Cys Leu Glu Cys Trp Ala Glu Asn Leu Gly Pro His His Thr Ser

1 Ser Leu Leu Asn Pro Arg His Leu Pro Ser Ile Pro Ala Met Phe Pro 15  
20 25 30  
5 Val Ser Ser Gly Cys Phe Gln Glu Gln Gln Met Asn Lys Ser Leu 45  
35 40 45  
10 Val Ser Cys Leu Phe Val Leu His Phe Val Leu His Cys Ile Phe Xaa 60  
50 55 60  
15 Met Leu Ser Thr Ser Glu Tyr Ser Gln Ser Pro Lys Met Glu Ser Leu 15  
1 5 10  
20 Ser Ser His Arg Ile Asp Glu Asp Gly Glu Asn Thr Gln Ile Glu Asp 30  
20 25 30  
30 Thr Glu Pro Met Ser Pro Val Leu Asn Ser Lys Phe Val Pro Ala Glu 45  
35 40 45  
35 Asn Asp Ser Ile Leu Met Asn Pro Ala Gln Asp Gly Glu Val Gln Leu 60  
50 55 60  
Ser Gln Asn Asp Asp Lys Thr Lys Gly Asp Asp Thr Asp Thr Arg Asp 80  
65 70 75  
40 Asp Ile Ser Ile Leu Ala Thr Gly Cys Lys Gly Arg Glu Glu Thr Val 95  
85 90 95  
45 Ala Glu Glu Val Cys Ile Asp Leu Thr Cys Asp Ser Gly Ser Gln Ala 110  
100 105 110  
Val Pro Ser Pro Ala Thr Arg Ser Glu Ala Leu Ser Ser Val Leu Asp 125  
115 120 125  
50 Gln Glu Glu Ala Met Glu Ile Lys Glu His Pro Glu Glu Gly Ser 140  
130 135 140  
Ser Gly Ser Glu Val Glu Glu Ile Pro Glu Thr Pro Cys Glu Ser Gln 160  
145 150 155 160  
55 Gly Glu Glu Leu Lys Glu Glu Asn Met Glu Ser Val Pro Leu His Leu 175  
165 170 175  
Ser Leu Thr Glu Thr Gln Ser Gln Gly Leu Cys Leu Arg Arg His Pro 190  
180 185 190  
60

550

551

Lys Lys Lys Lys  
195

5

(2) INFORMATION FOR SEQ ID NO: 325:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 252 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 325:

Met Gly Cys Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg Arg Lys  
1 5 10 15

Arg Leu Leu Glu Gln Glu-Lys Ser Leu Ala Gly Trp Ala Leu Val Leu  
20 25 30

Ala Xaa Xaa Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp  
35 40 45

Phe Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr  
50 55 60

Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val  
65 70 75 80

Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val  
85 90 95

Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val Ala Arg Lys  
100 105 110

Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp  
115 120 125

Ile Gln Thr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln  
130 135 140

Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala  
145 150 155 160

Ala Arg Xaa His Gln Arg Xaa Leu Leu Ala Ala Ile Asn Ala Phe Arg  
165 170 175

Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met  
180 185 190

Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn  
195 200 205

Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu Ala  
210 215 220

Gly Lys Leu Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro  
225 230 235 240

Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa  
245 250

5

(2) INFORMATION FOR SEQ ID NO: 326:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 68 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 326:

Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val  
1 5 10 15

Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala  
20 25 30

Val Ile Lys Asn Asn Ser His Tyr Gln Thr Ser Lys Ala Leu Glu Leu  
35 40 45

Glu Lys Thr Thr Glu Asn Lys Glu Ser Asn Pro Phe Ile Leu Gln Val  
50 55 60

Asn Lys Leu Xaa  
65

30

(2) INFORMATION FOR SEQ ID NO: 327:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 84 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 327:

Met Gly Glu Gly Lys Asn Gly Phe Gly Phe Val His Thr Ala Asp  
1 5 10 15

Ala Cys Trp Glu Gly Val His Ser Glu Pro Val Cys Arg Thr Val His  
20 25 30

Thr Val His Thr Cys His His Gln Ala Phe Leu Val Leu Ile Gly Trp  
35 40 45

Ser Lys Ser Gly Lys Glu Arg Lys Glu Ala Phe Leu Thr Ala Ile Ile  
50 55 60

Leu Asn Ser Arg Ser Ile His Ile Ser Cys Ser Trp Pro Pro Ser Pro  
65 70 75 80

Val Pro Gln Xaa

55

(2) INFORMATION FOR SEQ ID NO: 328:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 328:

5 Met Leu Leu Ile Asn Leu Leu Trp Leu Val Thr Met Ile Lys Ser Val 15  
10  
Ile Asn Asn Ile Ile Leu Phe Leu Lys Lys Lys Ser Leu Phe Phe 20  
25  
10 Ile Asp Ser Val 35

(2) INFORMATION FOR SEQ ID NO: 329:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 63 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 329:

20 Met Thr Phe Pro Phe Glu Lys Lys Ile Val Ala Phe Ser Ala Phe Tyr 15  
25  
Leu Ile Pro Gly Glu Ser Arg Leu Ala Pro Thr Phe Asn Pro Ser Ala 30  
35  
Asp Met Thr Val Ile Leu Arg Gly Arg Ala Gln His Lys Thr Ala Met 45  
Leu Glu Ser Tyr Asn Trp Lys Val Ser Cys Gln Leu Arg Glu Xaa 60  
50  
55

(2) INFORMATION FOR SEQ ID NO: 330:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 35 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 330:

40 Met His Ser Lys Gly Ser Ser Leu Leu Leu Phe Leu Pro Gln Leu Ile 15  
45  
Leu Ile Leu Pro Val Cys Ala His Leu His Glu Glu Leu Asn Cys Cys 20  
25  
Phe His Arg 35

(2) INFORMATION FOR SEQ ID NO: 331:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 331:

5 Met Gly Ala Leu Val Leu Leu Cys Leu Leu Val Gly Val Gln Gln 15  
10  
Ser Gly Ser Val Trp Asp Ser 20

(2) INFORMATION FOR SEQ ID NO: 332:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 332:

20 Met Gln Ser Ala Glu Ile Leu Ser Trp Thr Asp Val Leu His Asp Phe 15  
25  
Leu Phe Ser Leu Phe Leu Trp Pro Ala Phe Glu Asp Arg Ala Leu Leu 30  
Ile Phe Thr Leu Asn Gln Ile Val 40

(2) INFORMATION FOR SEQ ID NO: 333:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 111 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 333:

40 Met Gln Ser Leu Val Gln Trp Gly Leu Asp Ser Tyr Asp Tyr Leu Gln 15  
45  
Asn Ala Pro Pro Gly Phe Phe Pro Arg Leu Gly Val Ile Gly Phe Ala 20  
25  
Gly Leu Ile Gly Leu Leu Ala Arg Gly Ser Lys Ile Lys Lys Leu 35  
40  
Val Tyr Pro Pro Gly Phe Met Gly Leu Ala Ala Ser Leu Tyr Tyr Pro 50  
55  
Gln Gln Ala Ile Val Phe Ala Gln Val Ser Gly Glu Arg Leu Tyr Asp 65  
70  
Trp Gly Leu Arg Gly Tyr Ile Val Ile Glu Asp Leu Trp Lys Glu Asn 85  
90  
Phe Gln Lys Pro Gln Asn Val Lys Asn Ser Pro Gly Thr Lys Xaa 100  
105  
110

## (2) INFORMATION FOR SEQ ID NO: 314:

5 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 106 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 314:

10 Met Ala Pro Ser Leu Leu Leu Ala Pro Leu Cys Ser Leu Glu Ala  
 1 5 10 15

15 Val Leu Ser Ser Pro Leu Glu Lys Gln Cys Gln Leu Pro Gly Ile Phe.  
 20 25 30

Cys Gln Leu Gln Leu Pro Cys Pro Leu Leu Leu Ser Ala Gln Leu Leu  
 35 40 45

20 Lys Gly Ile Val Xaa Pro Arg Cys Pro Ala Ser Leu Pro Gln Pro Pro  
 50 55 60

His Pro Ala Pro Ser Trp His Leu Pro Leu His Cys Thr Glu Arg Xaa  
 65 70 75 80

25 Pro His His Leu Pro Leu Gln Gly Ser Ser Asn Met Glu Glu Xaa  
 85 90 95

30 Asn Tyr Arg Gly Tyr Xaa Asp Ala Gln Leu  
 100 105

## (2) INFORMATION FOR SEQ ID NO: 315:

35 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 50 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 315:

40 Met Thr Thr Cys Leu Phe Gly Leu Leu Ser Cys Glu Met Ser Ala Gln  
 1 5 10 15

45 Val Ser Gln Lys Ser Cys Val Tyr Asp Glu Ser Glu Cys Phe Ser Ser  
 20 25 30

Val Gly Gln Leu Leu Ala Leu Leu Ile Leu Val Tyr Val Leu Pro Ser  
 35 40 45

50 Ile Xaa  
 50

## (2) INFORMATION FOR SEQ ID NO: 316:

55 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 48 amino acids  
 (B) TYPE: amino acid

60

## (2) INFORMATION FOR SEQ ID NO: 316:

5 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 316:

10 Met Leu Trp Lys Cys Ser Gln Asn Ile Ala Arg Cys Leu Leu Leu  
 1 5 10 15

Leu Ala Leu Val Glu Ile Lys Leu Glu Asp Leu Gln Ser Gln Leu His  
 20 25 30

10 Pro Thr Trp Lys Ser Ile Pro Gly Pro Ser Pro Arg Asn Gln His Arg  
 35 40 45

15 (2) INFORMATION FOR SEQ ID NO: 317:

20 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 317:

25 Met Leu Ile Pro Leu Gln Cys Leu Phe Ser Ser Asp Arg Met Leu Thr  
 1 5 10 15

30 Phe Leu Thr Pro Trp Gln Lys Gly Glu Lys Cys Val Leu Gly Trp Val  
 20 25 30

Thr Lys Phe Leu Ser Glu Ile Ser Xaa  
 35 40

35 (2) INFORMATION FOR SEQ ID NO: 318:

40 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 76 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 318:

45 Met Thr Phe Ser Ser Leu Lys Leu Phe Val Leu Thr Cys Ile Ile Lys  
 1 5 10 15

Gly Leu Glu Arg Phe Ile Ile Leu Arg Glu Val Cys Asn Gln Glu Ile  
 20 25 30

Gln Arg Ser Leu Ser Ser Asn Leu Val His Val Leu Leu Gln Pro Ala  
 35 40 45

50 Thr Phe Lys Asp Val Leu Val Thr Glu Ile Ile Cys Leu Cys Met Cys  
 55 60 65

Leu Tyr Ser Ile Lys Tyr Met Pro Pro Gln Lys Lys  
 65 70 75

60

(2) INFORMATION FOR SEQ ID NO: 339:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 339:

10 Lys Val Tyr Ile Phe Leu Ile Phe Met Val Leu Ile Leu Pro Ser Leu 15

Gly Leu Thr Arg Tyr Met Pro Pro Xaa Ser Xaa Leu Asn Ser Glu 30

15

(2) INFORMATION FOR SEQ ID NO: 340:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 340:

20 Met Ala Lys Ile Ser Pro Phe Glu Val Val Lys Arg Thr Ser Val Pro 15

30 Val Leu Val Gly Leu Val Ile Val Ile Val Ala Thr Glu Leu Met Val 30

Pro Gly Thr Ala Ala Val Thr Gly Lys 40

35

(2) INFORMATION FOR SEQ ID NO: 341:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 341:

40 Met Arg Leu Phe Ile Gly Phe Leu Leu Phe Ser Phe Gly Leu 15

Leu Arg Gln Pro Ser Leu Ser Ala Glu His 25

50

(2) INFORMATION FOR SEQ ID NO: 342:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 26 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 342:

60

Met Val Phe Ser Val Ser Ala Leu Ala Leu Leu Met Leu Leu 15

5 Arg Ser Ser Asp Leu Ala Lys Lys Thr Glu 25

(2) INFORMATION FOR SEQ ID NO: 343:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 157 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 343:

10 Met Ser Leu Glu Phe Tyr Gln Lys Lys Ser Arg Trp Pro Phe Ser 15

20 Asp Glu Cys Ile Pro Trp Glu Val Trp Thr Val Lys Val His Val Val 30

35 Ala Leu Ala Thr Glu Gln Arg Gln Ile Cys Arg Glu Lys Val Gly 45

50 Glu Lys Leu Cys Glu Lys Ile Ile Asn Ile Val Glu Val Met Asn Arg 60

65 His Glu Tyr Leu Pro Lys Met Pro Thr Gln Ser Glu Val Asp Asn Val 80

85 Phe Asp Thr Gly Leu Arg Asp Val Gln Pro Tyr Leu Tyr Lys Ile Ser 95

100 Phe Gln Ile Thr Asp Ala Leu Gly Thr Ser Val Thr Thr Thr Met Arg 110

115 Arg Leu Ile Lys Asp Thr Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly 125

130 Ser Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu 140

145 Gly Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp Xaa 155

(2) INFORMATION FOR SEQ ID NO: 344:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 520 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 344:

1 Met Phe Leu Leu Pro Leu Pro Ala Ala Gly Arg Val Val Val Arg Arg 15

20 Leu Ala Val Arg Arg Phe Gly Ser Arg Ser Leu Ser Thr Ala Asp Met

20 25 30  
 Thr Lys Gly Leu Val Leu Gly Ile Tyr Ser Lys Glu Lys Glu Asp Asp  
 35 40 45  
 5 Val Pro Gln Phe Thr Ser Ala Gly Glu Asn Phe Asp Lys Leu Ala  
 50 55 60  
 Gly Lys Leu Arg Glu Thr Leu Asn Ile Ser Gly Pro Pro Leu Lys Ala  
 65 70 75 80  
 Gly Lys Thr Arg Thr Phe Tyr Gly Leu His Gln Asp Phe Pro Ser Val  
 85 90 95  
 15 Val Leu Val Gly Leu Gly Lys Lys Ala Ala Gly Ile Asp Glu Gln Glu  
 100 105 110  
 Asn Trp His Glu Gly Lys Glu Asn Ile Arg Ala Val Ala Ala Gly  
 115 120 125  
 20 Cys Arg Gln Ile Gln Asp Leu Glu Leu Ser Ser Val Glu Val Asp Pro  
 130 135 140  
 Cys Gly Asp Ala Gln Ala Ala Glu Gly Ala Val Leu Gly Leu Tyr  
 145 150 155 160  
 Glu Tyr Asp Asp Leu Lys Gln Lys Lys Lys Met Ala Val Ser Ala Lys  
 165 170 175  
 30 Leu Tyr Gly Ser Gly Asp Gln Glu Ala Trp Gln Lys Gly Val Leu Phe  
 180 185 190  
 Ala Ser Gly Gln Asn Leu Ala Arg Gln Leu Met Glu Thr Pro Ala Asn  
 195 200 205  
 35 Glu Met Thr Pro Thr Arg Phe Ala Glu Ile Ile Glu Lys Asn Leu Lys  
 210 215 220  
 Ser Ala Ser Ser Lys Thr Glu Val His Ile Arg Pro Lys Ser Trp Ile  
 225 230 235 240  
 Glu Glu Gln Ala Met Gly Ser Phe Leu Ser Val Ala Lys Gly Ser Asp  
 245 250 255  
 45 Glu Pro Pro Val Phe Leu Glu Ile His Tyr Lys Gly Ser Pro Asn Ala  
 260 265 270  
 Asn Glu Pro Pro Leu Val Phe Val Gly Lys Gly Ile Thr Phe Asp Ser  
 275 280 285  
 50 Gly Gly Ile Ser Ile Lys Ala Ser Ala Asn Met Asp Met Arg Ala  
 290 295 300  
 Asp Met Gly Gly Ala Ala Thr Ile Cys Ser Ala Ile Val Ser Ala Ala  
 305 310 315 320  
 Lys Leu Asn Leu Pro Ile Asn Ile Ile Gly Leu Ala Pro Leu Cys Glu  
 325 330 335  
 60 Asn Met Pro Ser Gly Lys Ala Asn Lys Pro Gly Asp Val Val Arg Ala

340 345 350  
 Lys Asn Gly Lys Thr Ile Gln Val Asp Asn Thr Asp Ala Glu Gly Arg  
 355 360 365  
 5 Leu Ile Leu Ala Asp Ala Leu Cys Tyr Ala His Thr Phe Asn Pro Lys  
 370 375 380  
 Xaa Ile Leu Asn Ala Thr Leu Thr Gly Ala Met Asp Val Ala Leu  
 385 390 395 400  
 Gly Ser Gly Ala Thr Gly Val Phe Thr Asn Ser Ser Trp Leu Trp Asn  
 405 410 415  
 15 Lys Leu Phe Glu Ala Ser Ile Glu Thr Gly Asp Arg Val Trp Arg Met  
 420 425 430  
 Pro Leu Phe Glu His Tyr Thr Arg Gln Val Val Asp Cys Gln Leu Ala  
 435 440 445  
 20 Asp Val Asn Asn Ile Gly Lys Tyr Arg Ser Ala Gly Ala Cys Thr Ala  
 450 455 460  
 Ala Ala Phe Leu Lys Glu Phe Val Thr His Pro Lys Trp Ala His Leu  
 465 470 475 480  
 Asp Ile Ala Gly Val Met Thr Asn Lys Asp Glu Val Pro Tyr Leu Arg  
 485 490 495  
 30 Lys Gly Met Thr Gly Arg Pro Thr Arg Thr Leu Ile Glu Phe Leu Leu  
 500 505 510  
 Arg Phe Ser Gln Asp Asn Ala Xaa  
 515 520  
 35  
 (2) INFORMATION FOR SEQ ID NO: 345:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 39 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 345:  
 45 Thr Ile Leu Phe Leu Phe Leu Gln Leu Ser Ala Leu Arg Leu Ile Val  
 1 5 10 15  
 Gly Lys Asp Ser Ile Asp Ile Asp Ile Ser Ser Arg Arg Arg Glu Asp  
 20 25 30  
 Gln Ser Leu Arg Leu Asn Ala  
 35  
 55  
 (2) INFORMATION FOR SEQ ID NO: 346:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 234 amino acids  
 60



(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 346:

5 Met Thr Ser Glu Leu Asp Ile Phe Val Gly Asn Thr Thr Leu Ile Asp  
1 5 10 15  
Glu Asp Val Tyr Arg Leu Trp Leu Asp Gly Tyr Ser Val Thr Asp Ala  
20 25 30  
10 Val Ala Leu Arg Val Arg Ser Gly Ile Leu Glu Gln Thr Gly Ala Thr  
35 40 45  
15 Ala Ala Val Leu Gln Ser Asp Thr Met Asp His Tyr Arg Thr Phe His  
50 55 60  
Met Leu Glu Arg Leu His Ala Pro Pro Lys Leu Leu His Gln Leu  
65 70 75 80  
20 Ile Phe Gln Ile Pro Pro Ser Arg Gln Ala Leu Leu Ile Glu Arg Tyr  
85 90 95  
Tyr Ala Phe Asp Glu Ala Phe Val Arg Glu Val Leu Gly Lys Lys Leu  
100 105 110  
25 Ser Lys Gly Thr Lys Lys Asp Leu Asp Ile Ser Thr Lys Thr Gly  
115 120 125  
30 Ile Thr Leu Lys Ser Cys Arg Arg Gln Phe Asp Asn Phe Lys Arg Val  
130 135 140  
Phe Lys Val Val Glu Glu Met Arg Gly Ser Leu Val Asp Asn Ile Gln  
145 150 155 160  
35 Gln His Phe Leu Leu Ser Asp Arg Leu Ala Arg Asp Tyr Ala Ala Ile  
165 170 175  
Val Phe Phe Ala Asn Asn Arg Phe Glu Thr Gly Lys Lys Lys Leu Gln  
180 185 190  
40 Tyr Leu Ser Phe Gly Asp Phe Ala Phe Cys Ala Glu Leu Met Ile Gln  
195 200 205  
45 Asn Trp Thr Leu Gly Pro Val Asp Ser Gln Met Asp Asp Met Asp Met  
210 215 220  
Asp Leu Asp Arg Asn Phe Ser Arg Thr Xaa  
225 230  
50

(2) INFORMATION FOR SEQ ID NO: 347:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 169 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 347:

60 Met Ala Ala Ala Val Ala Gly Met Leu Arg Gly Gly Leu Leu Pro Gln

1 5 10 15  
Ala Gly Arg Leu Pro Thr Thr Leu Gln Thr Val Arg Tyr Gly Ser Lys Ala  
20 25 30  
5 Val Thr Arg His Arg Arg Val Met His Phe Gln Arg Gln Lys Leu Met  
35 40 45  
10 Ala Val Thr Glu Tyr Ile Pro Pro Lys Pro Ala Ile His Pro Ser Cys  
50 55 60  
Leu Pro Ser Pro Ser Pro Gln Glu Ile Gly Leu Ile Arg  
65 70 75 80  
15 Leu Leu Arg Arg Glu Ile Ala Ala Val Phe Gln Asp Asn Arg Met Ile  
85 90 95  
Ala Val Cys Gln Asn Val Ala Leu Ser Ala Glu Asp Lys Leu Leu Ile  
100 105 110  
20 Ala Thr Pro Ala Ala Glu Thr Gln Asp Pro Asp Glu Gly Leu Pro Gln  
115 120 125  
25 Pro Gly Pro Glu Ser Pro Ser Trp Arg Ile Pro Ser Thr Lys Ile Cys  
130 135 140  
Cys Pro Phe Leu Trp Gly Thr Thr Cys Cys Trp Ser Val Lys Ser Pro  
145 150 155 160  
30 Arg Ser Arg Arg Trp Tyr Gly Ser Xaa  
165

35 (2) INFORMATION FOR SEQ ID NO: 348:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 348:

40 Met Lys Arg Ser Phe Leu Leu Pro Leu Leu Val Gly Phe Leu Asp  
1 5 10 15  
45 Thr Ala His Leu Ile Leu Leu Glu Thr Leu Ser Val Cys Leu Trp Leu  
20 25 30  
Pro Ser Leu Ile Asp Ser Arg Cys Val Met Ser  
35 40 45  
50

55 (2) INFORMATION FOR SEQ ID NO: 349:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 78 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 349:

60

Met Lys Glu Gly Pro Pro Cys Lys Arg His His Tyr Tyr Gln Asn Cys  
1 5 10 15

5 Gly Ala Lys Leu Leu Val Ser Leu Phe Gly Gly Thr Asn Gln Ile His  
20 25 30

Leu Leu Glu Thr Gln Val Gly Thr Glu Lys Gly Gly Arg Ile Trp  
35 40 45

10 Glu Glu Lys Trp Arg Ile Ser Thr Val Leu Phe Ile Ser Val Asn  
50 55 60

15 Ser Tyr Val Glu Gly Ser Val Leu Glu Ile Lys Leu Phe Tyr  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 350:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 350:

Met Ser Glu Ile Leu Ser Leu Leu Phe Cys Leu Leu Gly Pro Ala Leu  
1 5 10 15

30 Asp Glu Arg Arg Glu Lys Asp  
20

35 (2) INFORMATION FOR SEQ ID NO: 351:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 274 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 351:

Met Ser Ser Ala Gly Thr Ala Thr Pro Leu Glu Met Asp His Lys Leu  
1 5 10 15

40 Thr Ser Gln Pro Gly Arg Pro Ser Phe Tyr Cys Asn Ser Arg His Ser  
20 25 30

45 Ile Val Gly Ser Ser His Gln Leu Leu Gly Phe Trp Phe Ser His Leu Glu  
35 40 45

50 Ser Ser Gly Leu Lys Val Phe Gln Val Ser Leu Pro Cys Glu Cys Val  
50 55 60

55 Asn Leu Pro Thr Arg Ile Ala Ser Val Leu Ser Leu Met Ser Leu  
65 70 75 80

60 Leu Val Val Gly Gln Ala Pro Ala Trp Glu Gly Ser Leu Leu Arg Gly  
85 90 95

Arg Pro Ala Gly Gly Ala His Leu Cys Ala Met Xaa Val Ile Glu Gly  
100 105 110

5 Leu Val Val Asp Val Gly Glu Arg Ile Leu His Gly Gln Arg Glu Val  
115 120 125

Gly Gln Val Ser Gln Val Leu Pro Ala Leu Ser Leu Gly Leu Val Phe  
130 135 140

10 Leu Cys Gln Gly Thr Val Glu Lys Val Ser Gly Ala Ala His Cys Ser  
145 150 155 160

Ser Leu Leu Cys Cys Leu Pro Trp Gln Cys Ser Gly Gly Phe Pro  
165 170 175

15 Thr Xaa Arg Cys Ser Arg Pro Tyr Phe Ser Ser His Lys Gly Val Ala  
180 185 190

20 Ala Thr Leu Ala Leu Thr Cys His Cys Asp Lys Val His Val Ala Gly  
195 200 205

Leu Gly Lys Asp Trp Ala Ile Glu Gln Arg Arg Thr Cys Glu Ser  
210 215 220

25 Asp Xaa Glu Xaa Xaa Pro Phe Thr Leu Ala Gly Leu Val Leu Val Leu  
225 230 235 240

Arg Phe Cys Gln Val Val Leu Val Trp Ile Pro Gln Leu Gly Asp Lys  
245 250 255

30 His Trp Arg Gly Met Thr Arg Leu Gly Arg Val Ser Leu Thr Ser Ser  
260 265 270

35 Ile Xaa

(2) INFORMATION FOR SEQ ID NO: 352:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 352:

Met Ile Phe Thr Ser Val Thr Lys Gly Ile Leu Leu Ile Ala Leu Trp  
1 5 10 15

50 Val Pro Leu Phe His Phe Met Leu Ile Asp Ser Ile Leu Gly Pro Ser  
20 25 30

55 Arg Leu Leu Thr Asp Gly Val Pro Phe Asn Pro Trp His Val Xaa  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 353:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 353:

5

Met Lys Thr  
1

10

(2) INFORMATION FOR SEQ ID NO: 354:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 354:

Met Ser Ile Ser Gly Thr Asp Gly Leu Ile Leu Leu Leu Val Gly Leu  
1 5 10 15

Glu Ala Xaa Val Arg Ser Ser Lys Lys Trp Ile Pro Lys Ala Leu Xaa  
20 25 30

Val Thr Gln Ala Lys Trp Asn Ser Trp Pro Ser Arg Asn Ala Gly  
35 40 45

Phe Ala Leu His  
50

30

(2) INFORMATION FOR SEQ ID NO: 355:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 355:

Met Glu His Cys Leu Tyr His Ser Val His Gly Ile Asn Pro Tyr Ile  
1 5 10 15

His Lys Asn Thr His Pro Ser Ile Asn Ile Tyr Met Val Trp Asp Glu  
20 25 30

Gln Val Asn Ser Phe Glu Arg Glu Phe Val Pro Phe Phe Leu Ile  
35 40 45

Ile Leu Leu Asn Cys Cys Gln Leu Ser Asn Lys Gln Thr Glu Lys Leu  
50 55 60

Phe Gly Lys Thr Leu His Thr Pro Phe Leu Ser Ser Ala Leu Lys Tyr  
65 70 75

Arg Leu Asn Thr His Ile Leu Pro Val Phe Ser Tyr Ser Asp Ser Ile  
85 90 95

Leu Thr Cys His Leu Ile Leu Ala Ser Tyr Phe Ser His Val Tyr Leu  
100 105 110

60

Pro Val Thr Cys Ile Cys Tyr Leu Asn Arg Lys Lys Asn Ile Gln Lys  
115 120 125

Lys Lys Asn Xaa  
130

10

(2) INFORMATION FOR SEQ ID NO: 356:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 204 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 356:

Met Gly Ser Arg Asp His Leu Phe Lys Val Leu Val Val Gly Asp Ala  
1 5 10 15

Ala Val Gly Lys Thr Ser Leu Val Gln Asp Tyr Ser Gln Asn Ser Phe  
20 25 30

Ser Lys His Tyr Lys Ser Thr Val Gly Val Asp Phe Ala Leu Lys Val  
35 40 45

Leu Gln Trp Ser Asp Tyr Glu Ile Val Arg Leu Gln Leu Trp Asp Ile  
50 55 60

Ala Gly Gln Glu Arg Phe Thr Ser Met Thr Arg Leu Tyr Tyr Arg Asp  
65 70 75 80

Ala Ser Ala Cys Val Ile Met Phe Asp Val Thr Asn Ala Thr Thr Phe  
85 90 95

Ser Asn Ser Gln Arg Trp Lys Gln Asp Leu Asp Ser Lys Leu Thr Leu  
100 105 110

Pro Asn Gly Glu Pro Val Pro Cys Leu Leu Leu Ala Asn Lys Cys Asp  
115 120 125

Leu Ser Pro Trp Ala Val Ser Arg Asp Gln Ile Asp Arg Phe Ser Lys  
130 135 140

Glu Asn Gly Phe Thr Gly Trp Thr Glu Thr Ser Val Lys Glu Asn Lys  
145 150 155 160

Asn Ile Asn Glu Ala Met Arg Val Leu Ile Glu Lys Met Met Arg Asn  
165 170 175

Ser Thr Glu Asp Ile Met Ser Leu Ser Thr Gln Gly Asp Tyr Ile Asn  
180 185 190

Leu Gln Thr Lys Ser Ser Trp Ser Cys Xaa  
195 200

(2) INFORMATION FOR SEQ ID NO: 357:

60

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 357:

Met Ile Ser Leu Ile Phe Gln Leu Glu Glu Lys Leu Val Glu Lys  
1 5 10 15

10 Phe Phe Phe Leu Phe Phe Leu Lys Lys Gly Ser Gln Gly Ser  
20 25 30

Asn Leu Lys Ile Val Pro Arg His Met Arg Val Val Leu Arg Gly  
35 40 45

## (2) INFORMATION FOR SEQ ID NO: 358:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 73 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 358:

25 Met Thr Tyr Val Thr Cys Leu His Val Cys Leu Leu Val Glu Phe Leu  
1 5 10 15

30 Asn Ser Gln Leu Thr Asn His Arg Lys Tyr Tyr Phe Leu Ser Tyr Gly  
20 25 30

Phe Trp Phe Thr Gly Leu Arg Gly Phe Ser Glu Tyr Trp Trp Pro Gln  
35 40 45

35 Gln His Thr Ser Phe His Pro Asn Arg Asn Glu Ile Asn Phe Val Ser  
50 55 60

Thr Asp Asn Arg Ile Trp Val Thr Xaa  
65 70

## (2) INFORMATION FOR SEQ ID NO: 359:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 102 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 359:

50 Met Ser Asp Gln Glu Ala Lys Pro Ser Thr Glu Asp Leu Gly Asp Lys  
1 5 10 15

55 Lys Glu Gly Glu Tyr Ile Lys Leu Lys Val Ile Gly Gln Asp Ser Ser  
20 25 30

Glu Ile His Phe Lys Val Lys Met Thr Thr His Leu Lys Lys Leu Lys  
35 40 45

60 Glu Ser Tyr Cys Gln Arg Gln Gly Val Pro Met Asn Ser Leu Arg Phe  
85 90 95

50 55 60

Leu Phe Glu Gly Gln Arg Ile Ala Asp Asn His Thr Pro Lys Glu Leu  
65 70 75 80

5 Gly Met Glu Glu Asp Val Ile Glu Val Tyr Gln Glu Gln Thr Gly  
85 90 95

10 Gly His Ser Thr Val Xaa  
100

## (2) INFORMATION FOR SEQ ID NO: 360:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 48 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 360:

Met Gly Phe Pro Gln Trp His Leu Gly Asn His Ala Val Glu Pro Val  
1 5 10 15

25 Thr Ser Ile Leu Leu Leu Phe Leu Met Met Leu Gly Val Arg Gly  
20 25 30

30 Leu Leu Leu Val Gly Leu Val Tyr Leu Val Ser His Leu Ser Gln Arg  
35 40 45

## (2) INFORMATION FOR SEQ ID NO: 361:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 179 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 361:

45 Met Ser Ala Glu Val Lys Val Thr Gly Gln Asn Gln Glu Gln Phe Leu  
1 5 10 15

Leu Leu Ala Lys Ser Ala Lys Gly Ala Ala Leu Ala Thr Leu Ile His  
20 25 30

50 Gln Val Leu Glu Ala Pro Gly Val Tyr Val Phe Gly Glu Leu Leu Asp  
35 40 45

Met Pro Asn Val Arg Glu Leu Ala Glu Ser Asp Phe Ala Ser Thr Phe  
50 55 60

Arg Leu Leu Thr Val Phe Ala Tyr Gly Thr Tyr Ala Asp Tyr Leu Ala  
65 70 75 80

60 Glu Ala Arg Asn Leu Pro Leu Thr Glu Ala Gln Lys Asn Lys Leu  
85 90 95

Arg His Leu Ser Val Val Thr Leu Ala Ala Lys Val Lys Cys Ile Pro  
100 105 110

5 Tyr Ala Val Leu Leu Glu Ala Leu Ala Leu Arg Asn Val Arg Gln Leu  
115 120 125

Glu Asp Leu Val Ile Glu Ala Val Tyr Ala Asp Val Leu Arg Gly Ser  
130 135 140

10 Leu Asp Gln Arg Asn Gln Arg Leu Glu Val Asp Tyr Ser Ile Gly Arg  
145 150 155 160

Asp Ile Gln Arg Gln Asp Leu Ser Ala Ile Ala Arg Thr Leu Xaa Lys  
165 170 175

Asn His Xaa

(2) INFORMATION FOR SEQ ID NO: 362:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 362:

30 Met Lys Ser Ser Ser Leu Phe Phe Phe Leu Ala His Phe Ile His  
1 5 10 15

Ser His Asp Leu Pro Gly Leu Cys Arg  
20 25

(2) INFORMATION FOR SEQ ID NO: 363:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 224 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 363:

45 Met Lys Phe Ala Ala Ser Gly Xaa Phe Leu His His Met Ala Gly Leu  
1 5 10 15

Ser Ser Lys Leu Ser Met Ser Lys Ala Leu Pro Leu Thr Lys Val  
20 25 30

Val Gln Asn Asp Ala Tyr Thr Ala Pro Ala Leu Pro Ser Ile Arg  
35 40 45

55 Thr Lys Ala Leu Thr Asn Met Ser Arg Thr Leu Val Asn Lys Glu Glu  
50 55 60

Pro Pro Lys Glu Leu Pro Ala Ala Glu Pro Val Leu Ser Pro Leu Glu  
65 70 75 80

Gly Thr Lys Met Thr Val Asn Asn Leu His Pro Arg Val Thr Glu Glu  
85 90 95

5 Asp Ile Val Glu Leu Phe Cys Val Cys Gly Ala Leu Lys Arg Ala Arg  
100 105 110

Leu Val His Pro Gly Val Ala Glu Val Val Phe Val Lys Lys Asp Asp  
115 120 125

10 Ala Ile Thr Ala Tyr Lys Lys Tyr Asn Asn Arg Cys Leu Asp Gly Gln  
130 135 140

Pro Met Lys Cys Asn Leu His Met Asn Gly Asn Val Ile Thr Ser Asp  
145 150 155 160

15 Gln Pro Ile Leu Leu Arg Leu Ser Asp Ser Pro Ser Met Lys Lys Glu  
165 170 175

Ser Glu Leu Pro Arg Arg Val Asn Ser Ala Ser Ser Ser Asn Pro Pro  
180 185 190

Ala Glu Val Asp Pro Asp Thr Ile Leu Lys Ala Leu Phe Lys Ser Ser  
195 200 205

25 Gly Ala Ser Xaa Thr Thr Gln Pro Thr Glu Phe Lys Ile Lys Leu Xaa  
210 215 220

(2) INFORMATION FOR SEQ ID NO: 364:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 349 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 364:

40 Met Ser Lys Asn Cys Ile Lys Leu Leu Cys Glu Asp Pro Val Phe Ala  
1 5 10 15

Glu Tyr Ile Lys Cys Ile Leu Met Asp Glu Arg Thr Phe Leu Asn Asn  
20 25 30

Asn Ile Val Tyr Thr Phe Met Thr Thr Phe Leu Leu Lys Val Gln Ser  
35 40 45

50 Gln Val Phe Ser Glu Ala Asn Cys Ala Asn Leu Ile Ser Thr Leu Ile  
50 55 60

Thr Asn Leu Ile Ser Gln Tyr Gln Asn Leu Gln Ser Asp Phe Ser Asn  
65 70 75 80

Arg Val Glu Ile Ser Lys Ala Ser Ala Ser Leu Asn Gly Asp Leu Arg  
85 90 95

60 Ala Leu Ala Leu Leu Ser Val His Thr Pro Lys Gln Leu Asn Pro  
100 105 110

Ala Leu Ile Pro Thr Leu Gln Glu Leu Ser Lys Cys Arg Thr Cys  
115 120 125

5 Leu Gln Gln Arg Asn Ser Leu Gln Glu Gln Ala Lys Glu Arg Lys  
130 135 140

Thr Lys Asp Asp Glu Gly Ala Thr Pro Ile Lys Arg Arg Val Ser  
145 150 155 160

10 Ser Asp Glu Glu His Thr Val Asp Ser Cys Ile Ser Asp Met Lys Thr  
165 170 175

15 Glu Thr Arg Glu Val Leu Thr Pro Thr Ser Thr Ser Asp Asn Glu Thr  
180 185 190

Arg Asp Ser Ser Ile Ile Asp Pro Gly Thr Glu Gln Asp Leu Pro Ser  
195 200 205

20 Pro Glu Asn Ser Ser Val Lys Glu Tyr Arg Met Glu Val Pro Ser Ser  
210 215 220

Phe Ser Glu Asp Met Ser Asn Ile Arg Ser Gln His Ala Glu Gln  
225 230 235 240

25 Ser Asn Asn Gly Arg Tyr Asp Asp Cys Lys Glu Phe Lys Asp Leu His  
245 250 255

Cys Ser Lys Asp Ser Thr Leu Ala Glu Glu Ser Glu Phe Pro Ser  
260 265 270

30 Thr Ser Ile Ser Ala Val Leu Ser Asp Leu Ala Asp Leu Arg Ser Cys  
275 280 285

35 Asp Gly Gln Ala Leu Pro Ser Gln Asp Pro Glu Val Ala Leu Ser Leu  
290 295 300

Ser Cys Gly His Ser Arg Gly Leu Phe Ser His Met Gln Gln His Asp  
305 310 315 320

40 Ile Leu Asp Thr Leu Cys Arg Thr Ile Glu Ser Thr Ile His Val Val  
325 330 335

45 Thr Arg Ile Ser Gly Lys Gly Asn Gln Ala Ala Ser Xaa  
340 345

50 (2) INFORMATION FOR SEQ ID NO: 365:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 467 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

55 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 365:  
Met Leu His Gln Asp His Ile Thr Phe Ala Met Leu Leu Ala Arg Ile  
1 5 10 15

60 Lys Leu Lys Gly Thr Val Gly Glu Pro Thr Tyr Asp Ala Glu Phe Gln

20 25 30  
His Phe Leu Arg Gly Asn Glu Ile Val Leu Ser Ala Gly Ser Thr Pro  
35 40 45

5 Arg Ile Gln Gly Leu Thr Val Glu Gln Ala Glu Ala Val Val Arg Leu  
50 55 60

10 Ser Cys Leu Pro Ala Phe Lys Asp Leu Ile Ala Lys Val Gln Ala Asp  
65 70 75 80

Glu Gln Phe Gly Ile Thr Leu Asp Ser Ser Pro Glu Gln Thr Val  
85 90 95

15 Pro Tyr Leu Trp Ser Glu Glu Thr Pro Ala Thr Pro Ile Gly Gln Ala  
100 105 110

Ile His Arg Leu Leu Ile Gln Ala Phe Arg Pro Asp Arg Leu Leu  
115 120 125

20 Ala Met Ala His Met Phe Val Ser Thr Asn Leu Gly Glu Ser Phe Met  
130 135 140

Ser Ile Met Glu Gln Pro Leu Asp Leu Thr His Ile Val Xaa Thr Glu  
145 150 155 160

25 Val Lys Pro Asn Thr Pro Val Leu Met Cys Ser Val Pro Gly Tyr Asp  
165 170 175

30 Ala Ser Gly His Val Glu Asp Leu Ala Glu Gln Asn Thr Gln Ile  
180 185 190

Thr Ser Ile Ala Ile Gly Ser Ala Glu Gly Phe Asn Gln Ala Asp Lys  
195 200 205

35 Ala Ile Asn Thr Ala Val Lys Ser Gly Arg Trp Val Met Leu Lys Asn  
210 215 220

Val His Leu Ala Pro Gly Trp Leu Met Gln Leu Lys Lys Leu His  
225 230 235 240

Ser Leu Gln Pro His Ala Cys Phe Arg Leu Phe Leu Thr Met Glu Ile  
245 250 255

45 Asn Pro Lys Val Pro Val Asn Leu Leu Arg Ala Gly Arg Ile Phe Val  
260 265 270

Phe Glu Pro Pro Gly Xaa Lys Ala Asn Met Leu Arg Thr Phe Ser  
275 280 285

50 Ser Ile Pro Val Ser Arg Ile Cys Lys Ser Pro Asn Glu Arg Ala Arg  
290 295 300

Leu Tyr Phe Leu Leu Ala Trp Phe His Ala Ile Ile Gln Glu Arg Leu  
305 310 315 320

Arg Tyr Ala Pro Leu Gly Trp Ser Lys Lys Tyr Glu Phe Gly Glu Ser  
325 330 335

60 Asp Leu Arg Ser Xaa Cys Asp Thr Val Asp Thr Trp Leu Asp Asp Thr

Ala Lys Gly Arg Gln Asn Ile Ser Pro Asp Lys Ile Pro Trp Ser Ala  
355 360 365  
5 Leu Lys Thr Leu Met Ala Gln Ser Ile Tyr Gly Gly Arg Val Asp Asn  
370 375 380  
10 Glu Phe Asp Gln Arg Leu Leu Asn Thr Phe Leu Glu Arg Leu Phe Thr  
385 390 395  
Thr Arg Ser Phe Asp Ser Glu Phe Lys Leu Ala Cys Lys Val Asp Gly  
400 405 410 415  
15 His Lys Asp Ile Gln Met Pro Asp Gly Met Gln Ala Arg Gly Val Cys  
420 425 430  
Ala Val Gly Gly Val Ala Pro Arg His Pro Asp Ala Leu Leu Ala Gly  
435 440 445  
20 Pro Ala Gln Gln Arg Arg Glu Ser Pro Pro Tyr His Thr Gly Cys Gly  
450 455 460  
His Asp Gln  
465

## (2) INFORMATION FOR SEQ ID NO: 366:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 152 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 366:

Met Ala Asp Glu Ala Thr Arg Arg Val Val Ser Glu Ile Pro Val Leu  
1 5 10 15  
40 Lys Thr Asn Ala Gly Pro Arg Asp Arg Glu Leu Trp Val Gln Arg Leu  
20 25 30  
Lys Glu Glu Tyr Gln Ser Leu Ile Arg Tyr Val Glu Asn Asn Lys Asn  
35 40 45  
Ala Asp Asn Asp Trp Phe Arg Leu Glu Ser Asn Lys Glu Gly Thr Arg  
50 55 60  
50 Trp Phe Gly Lys Cys Trp Tyr Ile His Asp Leu Lys Tyr Glu Phe  
65 70 75 80  
Asp Ile Glu Phe Asp Ile Pro Ile Thr Tyr Pro Thr Thr Ala Pro Glu  
85 90 95  
55 Ile Ala Val Pro Glu Leu Asp Gly Lys Thr Ala Lys Met Tyr Arg Gly  
100 105 110  
Gly Lys Ile Cys Leu Thr Asp His Phe Lys Pro Leu Trp Gly Gln Glu  
115 120 125  
60

Cys Ala Gln Ile Trp Thr Ser Ser Ser His Gly Ser Gly Ala Gly Ser  
130 135 140  
5 Met Xaa Gly Ser Gly Asn Pro Xaa  
145 150  
10 (2) INFORMATION FOR SEQ ID NO: 367:  
(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 373 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 367:  
Met Tyr Asp Gly Thr Lys Glu Val Pro Met Asn Pro Val Lys Ile Tyr  
1 5 10 15  
20 Gln Val Cys Asp Ile Pro Gln Pro Gln Gly Ser Ile Ile Asn Pro Gly  
25 30  
Ser Thr Gly Ser Ala Pro Trp Asp Glu Lys Asp Asn Asp Val Asp Glu  
35 40 45  
Glu Asp Glu Glu Asp Glu Leu Asp Gln Ser Gln His His Val Pro Ile  
50 55 60  
30 Gln Asp Thr Phe Pro Phe Leu Asn Ile Asn Gly Ser Pro Met Ala Pro  
65 70 75 80  
Ala Ser Val Gly Asn Cys Ser Val Gly Asn Cys Ser Pro Glu Ala Val  
85 90 95  
35 Trp Pro Lys Thr Glu Pro Leu Glu Met Glu Val Pro Gln Ala Pro Ile  
100 105 110  
Gln Pro Phe Tyr Ser Ser Pro Glu Leu Trp Ile Ser Ser Leu Pro Met  
115 120 125  
40 Thr Asp Leu Asp Ile Lys Phe Gln Tyr Arg Gly Lys Glu Tyr Gly Gln  
130 135 140 145  
Thr Met Thr Val Ser Asn Pro Gln Gly Cys Arg Leu Phe Tyr Gly Asp  
145 150 155 160  
Leu Gly Pro Met Pro Asp Gln Glu Leu Phe Gly Pro Val Xaa Leu  
165 170 175  
50 Glu Gln Val Lys Phe Pro Gly Pro Glu His Ile Thr Asn Glu Lys Gln  
180 185 190  
Lys Leu Phe Thr Ser Lys Leu Leu Asp Val Met Asp Arg Gly Leu Ile  
195 200 205  
55 Leu Glu Val Ser Gly His Ala Ile Tyr Ala Ile Arg Leu Cys Gln Cys  
210 215 220 225  
Lys Val Tyr Trp Ser Gly Pro Cys Ala Pro Ser Leu Val Ala Pro Asn  
230 235 240

Leu Ile Glu Arg Gln Lys Lys Val Lys Leu Phe Cys Leu Glu Thr Phe  
245 250 255

5 Leu Ser Asp Leu Ile Ala His Gln Lys Gly Gln Ile Glu Lys Gln Pro  
260 265 270

Pro Phe Glu Ile Tyr Leu Cys Phe Gly Glu Glu Trp Pro Asp Gly Lys  
275 280 285

10 Pro Leu Glu Arg Lys Leu Ile Leu Val Gln Val Ile Pro Val Ala  
290 295 300

Arg Met Ile Tyr Glu Met Phe Ser Gly Asp Phe Thr Arg Ser Phe Asp  
305 310 315 320

Ser Gly Ser Val Arg Leu Gln Ile Ser Thr Pro Asp Ile Lys Asp Asn  
325 330 335

20 Ile Val Ala Gln Leu Lys Gln Leu Tyr Arg Ile Leu Gln Thr Gln Glu  
340 345 350

Ser Trp Gln Pro Met Gln Pro Thr Pro Ser Met Gln Leu Pro Pro Ala  
355 360 365

25 Leu Pro Pro Gln Xaa  
370

30 (2) INFORMATION FOR SEQ ID NO: 368:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 83 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 368:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15

Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 30

45 Trp Gly Phe Gly Arg Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Glu Thr Glu Pro Ala Thr Ser Ser  
50 55 60

50 His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80

Pro Asn Xaa

(2) INFORMATION FOR SEQ ID NO: 369:

60

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 369:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15

10 Tyr Trp Thr Met Xaa  
20

15 (2) INFORMATION FOR SEQ ID NO: 370:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 227 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 370:

Met Gly Ala Ser Ala Arg Leu Leu Arg Ala Val Ile Met Gly Ala Pro  
1 5 10 15

25 Gly Ser Gly Lys Gly Thr Val Ser Ser Arg Ile Thr Thr His Phe Glu  
20 25 30

30 Leu Lys His Leu Ser Ser Gly Asp Leu Leu Arg Asp Asn Met Leu Arg  
35 40 45

Gly Thr Glu Ile Gly Val Leu Ala Lys Ala Phe Ile Asp Gln Gly Lys  
50 55 60

35 Leu Ile Pro Asp Asp Val Met Thr Arg Leu Ala Leu His Glu Leu Lys  
65 70 75 80

Asn Leu Thr Gln Tyr Ser Trp Leu Leu Asp Gly Phe Pro Arg Thr Leu  
85 90 95

40 Pro Gln Ala Glu Ala Leu Asp Arg Ala Tyr Gln Ile Asp Thr Val Ile  
100 105 110

45 Asn Leu Asn Val Pro Phe Glu Val Ile Lys Gln Arg Leu Thr Ala Arg  
115 120 125

Trp Ile His Pro Ala Ser Gly Arg Val Tyr Asn Ile Glu Phe Asn Pro  
130 135 140

50 Pro Lys Thr Val Gly Ile Asp Asp Leu Thr Gly Glu Pro Leu Ile Gln  
145 150 155 160

Arg Glu Asp Asp Lys Pro Glu Thr Val Ile Lys Arg Leu Lys Ala Tyr  
165 170 175

Glu Asp Gln Thr Lys Pro Val Leu Glu Tyr Tyr Gln Lys Lys Gly Val  
180 185 190

60 Leu Glu Thr Phe Ser Gly Thr Glu Thr Asn Lys Ile Trp Pro Tyr Val  
195 200 205



Tyr Ala Phe Leu Gln Thr Lys Val Pro Gln Arg Ser Gln Lys Ala Ser  
210 215 220

5 Val Thr Pro  
225

10 (2) INFORMATION FOR SEQ ID NO: 371:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 79 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 371:

Met Phe Leu Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Leu Thr Gln  
1 5 10 15

Leu Lys Ile Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu  
20 25 30

25 Lys Ile Ile Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Thr Ser  
35 40 45

Thr Ala Ile Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asp Asn  
50 55 60

30 Lys Thr Ala Lys Gly Gly Gln Glu Ala Leu Thr Cys Thr Xaa  
65 70 75

35 (2) INFORMATION FOR SEQ ID NO: 372:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 372:

Phe Arg Ala Val Phe Pro Cys Cys Pro Phe Leu Thr Leu Met Leu Pro  
1 5 10 15

45 Leu Leu Glu Cys Leu Val Gly Met Ile Met Cys Tyr Leu Gly Ile Ser  
20 25 30

50 Phe Thr Asp Thr Arg Lys Thr Ala Gly Leu Lys Lys Lys Lys Lys  
35 40 45

Lys Xaa Xaa  
50

55

(2) INFORMATION FOR SEQ ID NO: 373:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 61 amino acids

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 373:

5 Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser  
1 5 10 15

10 Phe Ile Val Thr Pro Thr Ser Asn Thr Ala Ser Tyr Leu Trp  
20 25 30

Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Ser Trp Ala Cys  
35 40 45

15 Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu Xaa  
50 55 60

20 (2) INFORMATION FOR SEQ ID NO: 374:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 374:

Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe  
1 5 10 15

30 Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg  
20 25 30

35 Ile Leu Phe Phe Ile Val Phe Xaa  
35 40

(2) INFORMATION FOR SEQ ID NO: 375:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 44 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 375:

45 Met Cys Ser Gly Gln Ser Gln Val Trp Lys Met Ala Leu Gln Ala Leu  
1 5 10 15

50 Asp Ser Glu Thr Val Val Ile Leu Pro Asp Met His Leu Ile Leu Ser  
20 25 30

Leu Arg Leu Ile His Asn Ala Arg Pro Cys Leu Xaa  
35 40

(2) INFORMATION FOR SEQ ID NO: 376:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 203 amino acids

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 376:

5 Met Leu Ile Ser Glu Glu Ile Pro Phe Lys Asp Asp Pro Arg Asp 15  
1 5 10  
Glu Thr Tyr Lys Pro His Leu Glu Arg Glu Thr Pro Lys Pro Arg Arg 30  
20 25  
Lys Ser Gly Lys Val Lys Glu Glu Lys Glu Lys Lys Glu Ile Lys Val 45  
35 40  
Glu Val Glu Val Glu Val Lys Glu Glu Glu Asn Glu Ile Arg Glu Asp 60  
50 55 60  
Glu Glu Pro Arg Lys Arg Gly Arg Arg Arg Lys Asp Lys Ser 75 80  
65 70  
Pro Arg Leu Pro Lys Arg Arg Lys Lys Pro Pro Ile Gln Tyr Val Arg 95  
85 90  
Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His Pro Arg Tyr Leu 110  
100 105  
Gln His Ile Lys Tyr Gln His Leu Leu Lys Lys Tyr Val Cys 125  
115 120  
Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln Lys Gln Leu Leu 140  
130 135 140  
Arg His Ala Lys His His Thr Asp Gln Arg Asp Tyr Ile Cys Glu Tyr 155 160  
145 150  
Cys Ala Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met 175  
165 170  
Ile His Thr Gly Glu Lys His Tyr Asn Val Arg Ser Val Asp Leu Leu 190  
180 185  
Val Asp Lys Arg His Leu Leu Ile Gly Thr Xaa 200  
195 200

(2) INFORMATION FOR SEQ ID NO: 377:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 377:

50 Met Leu Pro Arg Arg Thr Phe Tyr Phe Thr Phe Ile Phe Ile Phe Phe 15  
1 5 10  
Leu Ala Ser Phe Thr Gly Phe Thr Leu Arg Ala Ser Phe 25  
20 25

(2) INFORMATION FOR SEQ ID NO: 378:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 136 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 378:

10 Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Met 15  
1 5 10  
Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val Leu 30  
20 25  
Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu Trp 45  
35 40  
Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg Met 60  
50 55  
Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val Gln 80  
65 70 75  
Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu Asn 95  
85 90  
Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met Pro Ser Phe 110  
100 105  
Phe Ser Cys His Leu Phe Cys Thr Leu Arg Trp Lys Tyr Phe Glu Val 125  
115 120  
Phe Tyr Asn His Lys Phe Leu Xaa 135  
130 135

(2) INFORMATION FOR SEQ ID NO: 379:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 41 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 379:

45 Met Ala Trp Arg Arg Arg Glu Pro Ala Ser Gly Leu Ala Ala Cys Trp 15  
1 5 10  
Leu Trp Arg Cys Ser Pro Trp Pro Cys Ala Cys Pro Gly Pro Gly Ala 30  
20 25  
Gly Leu Ser Ser Gly Ser Arg Pro Trp 40  
35 40

(2) INFORMATION FOR SEQ ID NO: 380:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 468 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 380:

5 Met Glu Phe Leu Lys Val Ala Arg Arg Lys Arg Glu Glu Lys Val 15  
 1 5 10 15  
 10 Gln Ile Gln Lys Glu Leu Ser Val Leu Glu Glu Asp Ile Lys Arg Val 30  
 20 25 30  
 Glu Glu Met Ser Gly Leu Tyr Ser Pro Val Ser Glu Asp Ser Thr Val 45  
 35 40 45  
 15 Pro Gln Phe Glu Ala Pro Ser Pro Ser His Ser Ser Ile Ile Asp Ser 60  
 50 55 60  
 Thr Glu Tyr Ser Gln Pro Pro Gly Phe Ser Gly Ser Ser Gln Thr Lys 80  
 65 70 75 80  
 20 Lys Gln Pro Trp Tyr Asn Ser Thr Leu Ala Ser Arg Arg Lys Arg Leu 95  
 85 90 95  
 Thr Ala His Phe Glu Asp Leu Glu Gln Cys Tyr Phe Ser Thr Arg Met 110  
 100 105 110  
 Ser Arg Ile Ser Asp Asp Ser Arg Thr Ala Ser Gln Leu Asp Glu Phe 125  
 115 120 125  
 30 Gln Glu Cys Leu Ser Lys Phe Thr Arg Tyr Asn Ser Val Arg Pro Leu 140  
 130 135 140  
 Ala Thr Leu Ser Tyr Ala Ser Asp Leu Tyr Asn Gly Ser Ser Ile Val 160  
 145 150 155 160  
 35 Ser Ser Ile Glu Phe Asp Arg Asp Cys Asp Tyr Phe Ala Ile Ala Gly 175  
 165 170 175  
 Val Thr Lys Lys Ile Lys Val Tyr Glu Tyr Asp Thr Val Ile Gln Asp 190  
 180 185 190  
 Ala Val Asp Ile His Tyr Pro Glu Asn Glu Met Thr Cys Asn Ser Lys 205  
 195 200 205  
 45 Ile Ser Cys Ile Ser Trp Ser Ser Tyr His Lys Asn Leu Leu Ala Ser 220  
 210 215 220  
 Ser Asp Tyr Glu Gly Thr Val Ile Leu Trp Asp Gly Phe Thr Gly Gln 240  
 225 230 235 240  
 50 Arg Ser Lys Val Tyr Gln Glu His Glu Lys Arg Cys Trp Ser Val Asp 255  
 245 250 255  
 Phe Asn Leu Met Asp Pro Lys Leu Leu Ala Ser Gly Ser Asp Asp Ala 270  
 260 265 270  
 Lys Val Lys Leu Trp Ser Thr Asn Leu Asp Asn Ser Val Ala Ser Ile 285  
 275 280 285  
 60 Glu Ala Lys Ala Asn Val Cys Cys Val Lys Phe Ser Pro Ser Ser Arg

(A) LENGTH: 29 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

5 Met Arg Lys Glu Asp Gly Phe Thr Phe Phe Phe Leu Phe Phe Phe 15  
 1 5 10 15  
 10 Val Val Gly Ser Lys Phe Val Asn Gly Asn Lys Leu Val 25  
 20 25 30  
 Asp Thr Val Lys Ser Val Leu Asp Lys Asp Arg Lys Glu Asp Asp Thr 430  
 420 425 430  
 Asn Glu Phe Val Ser Ala Val Cys Trp Arg Ala Leu Pro Asp Gly Glu 445  
 435 440 445  
 30 Ser Asn Val Leu Ile Ala Ala Asn Ser Gln Gly Thr Ile Lys Val Leu 460  
 450 455 460  
 Glu Leu Val Xaa 465  
 460 465  
 35 (2) INFORMATION FOR SEQ ID NO: 381:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 29 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 381:

45 Met Arg Lys Glu Asp Gly Phe Thr Phe Phe Phe Leu Phe Phe Phe 15  
 1 5 10 15  
 50 Val Val Gly Ser Lys Phe Val Asn Gly Asn Lys Leu Val 25  
 20 25 30  
 Asp Thr Val Lys Ser Val Leu Asp Lys Asp Arg Lys Glu Asp Asp Thr 430  
 420 425 430  
 Asn Glu Phe Val Ser Ala Val Cys Trp Arg Ala Leu Pro Asp Gly Glu 445  
 435 440 445  
 30 Ser Asn Val Leu Ile Ala Ala Asn Ser Gln Gly Thr Ile Lys Val Leu 460  
 450 455 460  
 Glu Leu Val Xaa 465  
 460 465  
 35 (2) INFORMATION FOR SEQ ID NO: 382:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 29 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 382:

Met Pro Leu Ala Pro Tyr Cys Asp Leu Leu Val Ala Leu Ser Phe Ala  
1 10 15

5 Leu Val Leu Glu Ser Pro Val Asp Ser Ser Asp Phe Thr  
20 25

10 (2) INFORMATION FOR SEQ ID NO: 383:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 138 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 383:

Met Asn Ser Leu Val Ser Trp Gln Leu Leu Leu Phe Leu Cys Ala Thr  
1 5 10 15

20 His Phe Gly Glu Pro Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg  
25 30

25 Pro Thr Gly Gln Gln Leu Glu Ser Leu Gly Leu Leu Ala Pro Gly Glu  
35 40 45

Gln Ser Leu Pro Cys Thr Glu Arg Lys Pro Ala Ala Thr Ala Arg Leu  
50 55 60

30 Ser Arg Arg Gly Thr Ser Leu Ser Pro Pro Glu Ser Ser Gly Ser  
65 70 75 80

Pro Gln Gln Pro Gly Leu Ser Ala Pro His Ser Arg Gln Ile Pro Ala  
85 90 95

35 Pro Gln Gly Ala Val Leu Val Gln Arg Glu Lys Asp Leu Pro Asn Tyr  
100 105 110

40 Asn Trp Asn Ser Phe Gly Leu Arg Phe Gly Lys Arg Glu Ala Ala Pro  
115 120 125

Gly Asn His Gly Arg Ser Ala Gly Arg Gly  
130 135

(2) INFORMATION FOR SEQ ID NO: 384:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 384:

Met Ser Cys Phe Ile Asp Ser Xaa Asp Ser Lys Ile Leu His Leu Leu  
1 5 10 15

Val Val Ser Phe Ile Cys Xaa Phe Leu Leu Ile Leu Thr His Gly  
20 25 30

Ile Leu Ile Leu Arg Xaa Phe Ser Val Xaa Xaa His Ser Leu Lys  
35 40 45

5 Asn Asn Leu Glu Glu Tyr Leu Ile Leu Met Asn Lys Ala Leu Leu Thr  
50 55 60

Arg Glu Asp Phe Phe Val Leu Pro Xaa Ala  
65 70

10

(2) INFORMATION FOR SEQ ID NO: 385:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 521 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 385:

20 Met Ser Ala Gly Glu Val Glu Arg Leu Val Ser Glu Leu Ser Gly Gly  
1 5 10 15

Thr Gly Gly Asp Glu Glu Glu Trp Leu Tyr Gly Asp Glu Asn Glu  
20 25 30

25 Val Glu Arg Pro Glu Glu Asn Ala Ser Ala Asn Pro Pro Ser Gly  
35 40 45

30 Ile Glu Asp Glu Thr Ala Glu Asn Gly Val Pro Lys Pro Lys Val Thr  
50 55 60

Glu Thr Glu Asp Asp Ser Asp Ser Asp Asp Asp Glu Asp Asp  
65 70 75 80

35 Val His Val Thr Ile Gly Asp Ile Lys Thr Gly Ala Pro Gln Tyr Gly  
85 90 95

Ser Tyr Gly Thr Ala Pro Val Asn Leu Asn Ile Lys Thr Gly Gly Arg  
100 105 110

40 Val Tyr Gly Thr Thr Gly Thr Lys Val Lys Gly Val Asp Leu Asp Ala  
115 120 125

45 Pro Gly Ser Ile Asn Gly Val Pro Leu Leu Glu Val Asp Leu Asp Ser  
130 135 140

Phe Glu Asp Lys Pro Trp Arg Lys Pro Gly Ala Asp Leu Ser Asp Tyr  
145 150 155 160

50 Phe Asn Tyr Gly Phe Asn Glu Asp Thr Trp Lys Ala Tyr Cys Glu Lys  
165 170 175

Gln Lys Arg Ile Arg Met Gly Leu Glu Val Ile Pro Val Thr Ser Thr  
180 185 190

55 Thr Asn Lys Ile Thr Val Gln Gln Gly Arg Thr Gly Asn Ser Glu Lys  
195 200 205

60 Glu Thr Ala Leu Pro Ser Thr Lys Ala Glu Phe Thr Ser Pro Pro Ser  
210 215 220

Leu Phe Lys Thr Gly Leu Pro Pro Ser Arg Arg Leu Pro Gly Ala Ile  
225 230 235  
5 Asp Val Ile Gly Gln Thr Ile Thr Ile Ser Arg Val Glu Gly Arg  
245 250 255  
Arg Ala Asn Glu Asn Ser Asn Ile Gln Val Leu Ser Glu Arg Ser Ala  
260 265 270  
10 Thr Glu Val Asp Asn Asn Phe Ser Lys Pro Pro Phe Phe Pro  
275 280 285  
Gly Ala Pro Thr His Leu Pro Pro Phe Leu Pro Pro  
290 295 300  
15 Pro Thr Val Ser Thr Ala Pro Pro Leu Ile Pro Pro Gly Phe Pro  
305 310 315 320  
Pro Pro Pro Gly Ala Pro Pro Ser Leu Ile Pro Thr Ile Glu Ser  
325 330 335  
20 Gly His Ser Ser Gly Tyr Asp Ser Arg Ser Ala Arg Ala Phe Pro Tyr  
340 345 350  
Gly Asn Val Ala Phe Pro His Leu Pro Gly Ser Ala Pro Ser Trp Pro  
355 360 365  
30 Ser Leu Val Asp Thr Ser Lys Gln Trp Asp Tyr Tyr Ala Arg Arg Glu  
370 375 380  
Lys Asp Arg Asp Arg Glu Arg Asp Arg Glu Arg Asp Arg Asp  
385 390 395 400  
35 Arg Asp Arg Glu Arg Glu Arg Thr Arg Glu Arg Glu Arg Glu Arg Asp  
405 410 415  
His Ser Pro Thr Pro Ser Val Phe Asn Ser Asp Glu Arg Tyr Arg  
420 425 430  
40 Tyr Arg Glu Tyr Ala Glu Arg Gly Tyr Glu Arg His Arg Ala Ser Arg  
435 440 445  
Glu Lys Glu Glu Arg His Arg Glu Arg Arg His Arg Glu Lys Glu Glu  
450 455 460  
45 Thr Arg His Lys Ser Ser Arg Ser Asn Ser Arg Arg Arg His Glu Ser  
465 470 475 480  
50 Glu Glu Gly Asp Ser His Arg Arg His Lys His Lys Lys Ser Lys Arg  
485 490 495  
Ser Lys Glu Gly Lys Glu Ala Gly Ser Glu Pro Ala Pro Glu Gln Glu  
500 505 510  
55 Ser Thr Glu Ala Thr Pro Ala Glu Xaa  
515 520 525  
60 Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe  
530 535 540 545 550 555 560 565 570 575 580 585 590 595 600 605 610 615 620 625 630 635 640 645 650 655 660 665 670 675 680 685 690 695 700 705 710 715 720 725 730 735 740 745 750 755 760 765 770 775 780 785 790 795 800 805 810 815 820 825 830 835 840 845 850 855 860 865 870 875 880 885 890 895 900 905 910 915 920 925 930 935 940 945 950 955 960 965 970 975 980 985 990 995

(2) INFORMATION FOR SEQ ID NO: 386:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 137 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 386:  
Met Asn Ser Arg Gly Ile Trp Leu Ala Tyr Ile Ile Leu Val Gly Leu  
1 5 10 15  
Leu His Met Val Leu Leu Ser Ile Pro Phe Phe Ser Ile Pro Val Val  
20 25 30  
15 Trp Thr Leu Thr Asn Val Ile His Asn Leu Ala Thr Tyr Val Phe Leu  
35 40 45  
His Thr Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala  
50 55 60  
20 Arg Leu Leu Thr His Trp Glu Gln Met Asp Tyr Gly Leu Gln Phe Thr  
65 70 75 80  
Ser Ser Arg Lys Phe Leu Ser Ile Ser Pro Ile Val Leu Tyr Leu Leu  
85 90 95  
Ala Ser Phe Tyr Thr Lys Tyr Asp Ala Ala His Phe Leu Ile Asn Thr  
100 105 110  
30 Ala Ser Leu Leu Ser Val Leu Leu Pro Lys Leu Pro Gln Phe His Gly  
115 120 125  
Val Arg Val Phe Gly Ile Asn Lys Tyr  
130 135  
35  
(2) INFORMATION FOR SEQ ID NO: 387:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 186 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 387:  
Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu  
1 5 10 15  
Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg  
20 25 30  
Glu Trp Leu Glu Arg Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe  
35 40 45  
55 Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Lys Cys  
50 55 60  
Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe  
65 70 75 80 85

Val Phe Leu Gly Ile Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu 85 95  
 5 Leu Val Ala Leu Ala Val Phe Gly Ala Cys Tyr Ile Leu Tyr Leu 100 110  
 Arg Thr Leu Gly Ser Lys Leu Val Leu Phe Gly Arg Gly Val Ser Pro 115 125  
 10 Ala His Gln Tyr Ala Leu Ala Gly Gly Ile Ser Phe Pro Phe Thr 130 140  
 Leu Ala Gly Ala Gly Ser Ala Val Phe Thr Val Leu Gly Ala Thr Leu 145 155  
 15 Val Val Ile Gly Ser His Ala Phe His Gln Ile Gly Ala Val Asp 160 175  
 Gly Gly Gly Leu Gln Met Gly Pro Val Xaa 180 185

(2) INFORMATION FOR SEQ ID NO: 388:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 388:

Met  
1

(2) INFORMATION FOR SEQ ID NO: 389:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 299 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 389:

Met Leu Ser Ile Phe Tyr Phe Ala Ile Pro Val Gly Ser Gly Leu Gly 1 15  
 Tyr Ile Ala Gly Ser Lys Val Lys Asp Met Ala Gly Asp Tyr His Tyr 20 30  
 Ala Leu Arg Val Thr Pro Gly Leu Gly Val Val Ala Val Leu Leu Leu 35 45  
 Phe Leu Val Val Arg Gly Pro Pro Arg Gly Ala Val Gly His Ser 50 60  
 Asp Leu Pro Pro Leu Asn Pro Thr Ser Tyr Trp Ala Asp Leu Arg Ala 65 75  
 60 Leu Ala Arg Asn Pro Ser Phe Val Leu Ser Ser Leu Gly Phe Thr Ala

Val Ala Phe Val Thr Gly Ser Leu Ala Leu Trp Ala Pro Ala Phe Leu 85 95  
 5 Leu Arg Ser Arg Val Val Leu Gly Glu Thr Pro Pro Cys Leu Pro Gly 100 110  
 Asp Ser Cys Ser Ser Asp Ser Leu Ile Phe Gly Leu Ile Thr Cys 115 125  
 Leu Thr Gly Val Leu Gly Val Gly Leu Gly Val Glu Ile Ser Arg Arg 130 140  
 10 Leu Arg His Ser Asn Pro Arg Ala Asp Pro Leu Val Cys Ala Thr Gly 145 155  
 Leu Leu Gly Ser Ala Pro Phe Leu Ser Leu Ala Cys Ala Arg 160 175  
 20 Gly Ser Ile Val Ala Thr Tyr Ile Phe Ile Phe Ile Gly Glu Thr Leu 180 190  
 Leu Ser Met Asn Trp Ala Ile Val Ala Asp Ile Leu Leu Tyr Val Val 195 205  
 25 Ile Pro Thr Arg Arg Ser Thr Ala Glu Ala Phe Gln Ile Val Leu Ser 210 220  
 His Leu Leu Gly Asp Ala Gly Ser Pro Tyr Leu Ile Gly Leu Ile Ser 225 235  
 30 Asp Arg Leu Arg Arg Asn Trp Pro Ser Phe Leu Ser Glu Phe Arg 240 255  
 Ala Leu Gln Phe Ser Leu Met Leu Cys Ala Phe Val Gly Ala Leu Gly 260 270  
 35 Gly Ala Leu Pro Gly His Arg His Leu His Xaa 275 285  
 (2) INFORMATION FOR SEQ ID NO: 390:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 49 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 390:

Met Gly Pro Gln Gly Trp Val Arg Pro Leu Lys Thr Ala Pro Lys Leu 1 15  
 Gly Glu Ala Ile Arg Leu Ile Leu Phe Leu Asn Phe Val Lys Gln Cys 20 30  
 Ile Ala Ser Val Asn Leu Cys Ile Leu Arg Leu Asn Ile Thr Pro Leu 35 45

Leu

- 5 (2) INFORMATION FOR SEQ ID NO: 391:
- 10 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 61 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 391:
- 15 Met Tyr Val Asn Tyr Gly Thr Arg Asn Tyr Ser Thr Glu Gly Pro Ala 15  
 Ala Leu Leu Asp Gln Ala Lys Leu Ser Leu Leu Val Trp Val Leu Cys 30  
 Phe Val Leu Leu Phe Val Cys Phe Cys Gly Leu Ser Tyr Val Val Ile 45  
 Ala Gln Val Pro Val Gly Leu Leu Cys Ile Thr Glu Xaa 60
- 20 (2) INFORMATION FOR SEQ ID NO: 392:
- 30 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 79 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 392:
- 35 Met Leu Trp Phe Ala Asn Phe Phe Thr Tyr Leu Phe Leu Ser Gln Ser 15  
 Val Ala Phe Val His Ile Ser His Ile Gly Val Arg Gln Val Asn Thr 30  
 Asn Cys Tyr Phe Ser Arg Lys Ser Tyr Cys Tyr Gly Ile Leu Asn Pro 45  
 Ile Asn Cys Ile Lys Gly Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys 60  
 Lys Lys Lys Lys Ile Pro Ala Gly Arg Xaa Leu Phe Pro Phe Gly 75
- 40 (2) INFORMATION FOR SEQ ID NO: 393:
- 55 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 36 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 393:
- 60

- Met Pro Gly Ala Phe Ser Glu Thr Val Ile Asn Asp Leu Leu Ser Leu 15  
 Phe Leu Val Leu Pro Ala Glu Ser Tyr Ser Thr Leu Ser Gly Val 30  
 Tyr Arg Asn Ala 35
- 10 (2) INFORMATION FOR SEQ ID NO: 394:
- 15 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 180 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 394:
- 20 Met Ala Gln Ser Arg Asp Gly Gly Asn Pro Phe Ala Glu Pro Ser Glu 15  
 Leu Asp Asn Pro Phe Gln Asp Pro Ala Val Ile Gln His Arg Pro Ser 30  
 Arg Gln Tyr Ala Thr Leu Asp Val Tyr Asn Pro Phe Glu Thr Arg Glu 45  
 Pro Pro Pro Ala Tyr Glu Pro Pro Ala Pro Ala Pro Leu Pro Pro Pro 60  
 Ser Ala Pro Ser Leu Gln Pro Ser Arg Lys Leu Ser Pro Thr Glu Pro 80  
 Lys Asn Tyr Gly Ser Tyr Ser Thr Gln Ala Ser Ala Ala Ala Thr 95  
 Ala Glu Leu Leu Lys Lys Gln Glu Glu Leu Asn Arg Lys Ala Glu Glu 110  
 Leu Asp Arg Arg Ser Glu Ser Cys Ser Met Leu Pro Trp Xaa Ala Gln 125  
 Leu Leu Asp Arg Thr Ile Gly Pro Leu Tyr Leu Leu Phe Val Gln Phe 140  
 Ser Pro Ala Phe Ser Arg Thr Ser Pro Trp Arg Ser Pro Lys Asn Phe 155  
 Arg Arg Leu Tyr Pro Cys Thr Thr Ser Gly Cys Ala Ala Arg Trp 175  
 Xaa Phe Ser Xaa 180
- 40 (2) INFORMATION FOR SEQ ID NO: 395:
- 50 (1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 395:
- 5 Met Pro Thr Pro Cys Thr Ser Leu Pro Ser Cys Cys Gln His Arg Ser  
1 5 10 15  
Ile Thr Met Thr Leu  
20
- 10
- (2) INFORMATION FOR SEQ ID NO: 396:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 60 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 396:
- 20 Met Pro Leu Phe Ile Pro Leu Ile Phe Phe Leu Ser Leu Leu His Cys  
1 5 10 15  
Gln Ser Lys His Pro Ile Gln Met Ser Leu Cys Met Cys Val Asn Ile  
20 25  
Ser Leu Val Trp Ser Pro Val Arg Trp Ile Phe Gly Ser Lys Gly Leu  
30 35 40 45  
Phe Ser Val His Leu Gln Ser Ser Gln Arg Pro Ser  
50 55 60
- 35
- (2) INFORMATION FOR SEQ ID NO: 397:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 152 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 397:
- 40 Met Ala Gly Pro Arg Pro Xaa Trp Arg Asp Gln Leu Leu Phe Met Ser  
1 5 10 15  
Ile Ile Val Leu Val Ile Val Ile Cys Leu Met Leu Tyr Ala Leu  
20 25 30  
Leu Trp Glu Ala Gly Asn Leu Thr Asp Leu Pro Asn Leu Arg Ile Gly  
35 40 45  
Phe Tyr Asn Phe Cys Leu Trp Asn Glu Asp Thr Ser Thr Leu Gln Cys  
50 55 60  
His Gln Phe Pro Glu Leu Glu Ala Leu Gly Val Pro Arg Val Gly Leu  
65 70 75 80  
Gly Leu Ala Arg Leu Gly Val Tyr Gly Ser Leu Val Leu Thr Leu Phe  
85 90 95
- 60

- Ala Pro Gln Pro Leu Leu Ala Gln Cys Asn Xaa Asp Glu Arg Ala  
100 105 110  
5 Trp Arg Leu Ala Val Gly Phe Leu Ala Val Ser Ser Val Leu Leu Ala  
115 120 125  
Gly Gly Leu Gly Leu Phe Leu Ser Tyr Val Trp Asn Gly Ser Xaa Ser  
130 135 140  
10 Pro Ser Arg Gly Leu Gly Phe Xaa  
145 150
- 15
- (2) INFORMATION FOR SEQ ID NO: 398:
- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 480 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 398:
- 20 Met Ser Asp Gly Phe Asp Arg Ala Pro Gly Ala Gly Arg Gly Arg Xaa  
1 5 10 15  
Arg Gly Leu Gly Arg Gly Gly Gly Pro Xaa Gly Gly Gly Phe Pro  
20 25 30  
30 Xaa Gly Xaa Xaa Pro Ala Glu Arg Xaa Arg His Gln Pro Pro Gln Pro  
35 40 45  
Lys Ala Pro Gly Phe Leu Gln Pro Xaa Pro Leu Arg Gln Pro Arg Thr  
50 55 60  
35 Thr Pro Pro Gly Ala Gln Cys Glu Val Pro Ala Ser Pro Gln Arg  
65 70 75 80  
40 Pro Ser Arg Pro Gly Ala Leu Pro Glu Gln Thr Arg Pro Leu Arg Ala  
85 90 95  
Pro Pro Ser Ser Gln Asp Lys Ile Pro Gln Gln Asn Ser Glu Ser Ala  
100 105 110  
45 Met Ala Lys Pro Gln Val Val Val Ala Pro Val Leu Met Ser Lys Leu  
115 120 125  
Ser Val Asn Ala Pro Glu Phe Tyr Pro Ser Gly Tyr Ser Ser Tyr  
130 135 140  
50 Thr Glu Ser Tyr Glu Asp Gly Cys Glu Asp Tyr Pro Thr Leu Ser Glu  
145 150 155 160  
Tyr Val Gln Asp Phe Leu Asn His Leu Thr Glu Gln Pro Gly Ser Phe  
165 170 175  
Glu Thr Glu Ile Glu Gln Phe Ala Glu Thr Leu Asn Gly Cys Val Thr  
180 185 190  
60 Thr Asp Asp Ala Leu Gln Glu Leu Val Glu Leu Ile Tyr Gln Gln Ala



195 200 205  
 Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr 210 220  
 5 Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln 225 235 240  
 Leu Leu Gln Arg Cys Arg Thr Glu Thr Glu Val Lys Asp Gln Ala 245 250 255  
 10 Ala Lys Gly Asp Glu Val Thr Arg Lys Arg Phe His Ala Phe Val Leu 260 265 270 275  
 Phe Leu Gly Glu Leu Tyr Leu Asn Leu Glu Ile Lys Gly Thr Asn Gly 275 280 285 288  
 Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Glu Leu Leu 290 295 300 303  
 20 Asn Ala Leu Phe Ser Asn Pro Met Asp Asn Leu Ile Cys Ala Val 305 310 315 320  
 Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Glu Asp Ala Trp Lys Glu 325 330 335 338  
 25 Lys Gly Lys Met Asp Met Glu Glu Ile Ile Gln Arg Ile Glu Asn Val 340 345 350 353  
 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Leu Lys 355 360 365 368  
 30 Leu Val Glu Leu Arg Ser Asn Trp Gly Arg Val His Ala Thr Ser 370 375 380 383  
 Thr Tyr Arg Glu Ala Thr Pro Glu Asn Asp Pro Asn Tyr Phe Met Asn 385 390 395 400  
 40 Glu Pro Thr Phe Tyr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp 405 410 415 418  
 Pro Asp Tyr Gln Glu Lys Tyr Gln Glu Leu Leu Glu Arg Glu Asp Phe 420 425 430 433  
 45 Phe Pro Asp Tyr Glu Glu Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp 435 440 445 448  
 Pro Tyr Leu Asp Asp Ile Asp Asp Glu Met Asp Pro Glu Ile Glu Glu 450 455 460 463  
 50 Ala Tyr Glu Lys Phe Cys Leu Glu Ser Glu Arg Lys Arg Lys Gln Xaa 465 470 475 480  
 55  
 60 (2) INFORMATION FOR SEQ ID NO: 399:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 423 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (X1) SEQUENCE DESCRIPTION: SEQ ID NO: 399:  
 5 Met Glu Pro Lys Thr Ile Thr Asp Ala Leu Ala Ser Ser Ile Ile Lys 15  
 1 Ser Val Leu Pro Asn Phe Leu Pro Tyr Asn Val Met Leu Tyr Ser Asp 20 25 30  
 Ala Pro Val Ser Glu Leu Ser Glu Leu Leu Leu Glu Val Val 35 40 45  
 15 Leu Pro Ala Leu Leu Glu Gln Gly His Thr Arg Gln Trp Leu Lys Gly 50 55 60  
 Leu Val Arg Ala Trp Thr Val Thr Ala Gly Tyr Leu Leu Asp Leu His 65 70 75 80  
 Ser Tyr Leu Leu Gly Asp Gln Glu Asn Glu Asn Ser Ala Asn Gln 85 90 95  
 25 Gln Val Asn Asn Gln His Ala Arg Asn Asn Asn Ala Ile Pro Val 100 105 110 113  
 Val Gly Glu Gly Leu His Ala Ala His Gln Ala Ile Leu Gln Gln Gly 115 120 125 128  
 30 Gly Pro Val Gly Phe Gln Xaa Tyr Arg Arg Pro Leu Asn Phe Pro Leu 130 135 140 143  
 Arg Ile Phe Leu Leu Ile Val Phe Met Cys Ile Thr Leu Leu Ile Ala 145 150 155 160  
 Ser Leu Ile Cys Leu Thr Leu Pro Val Phe Ala Gly Arg Trp Leu Met 165 170 175 178  
 40 Ser Phe Thr Thr Gly Thr Ala Lys Ile His Glu Leu Tyr Thr Ala Ala 180 185 190 193  
 Cys Gly Leu Tyr Val Cys Trp Leu Thr Ile Arg Ala Val Thr Val Met 195 200 205 208  
 45 Val Ala Trp Met Pro Gln Gly Arg Arg Val Ile Phe Gln Lys Val Lys 210 215 220 223  
 Glu Trp Ser Leu Met Ile Met Lys Thr Leu Ile Val Ala Val Leu Leu 225 230 235 240  
 50 Ala Gly Val Val Pro Leu Leu Leu Gly Leu Leu Phe Glu Leu Val Ile 245 250 255 258  
 Val Ala Pro Leu Arg Val Pro Leu Asp Gln Thr Pro Leu Phe Tyr Pro 260 265 270 273  
 Trp Gln Asp Trp Ala Leu Gly Val Leu His Ala Lys Ile Ile Ala Ala 275 280 285 288  
 60

Ile Thr Leu Met Gly Pro Gln Trp Trp Leu Lys Thr Val Ile Glu Gln  
290 295 300

Val Tyr Ala Asn Gly Ile Arg Asn Ile Asp Leu His Tyr Ile Val Arg  
305 310 315 320

Lys Leu Ala Ala Pro Val Ile Ser Val Leu Leu Leu Ser Leu Cys Val  
325 330 335

Pro Tyr Val Ile Ala Ser Gly Val Val Pro Leu Leu Gly Val Thr Ala  
340 345 350

Glu Met Gln Asn Leu Val His Arg Arg Ile Tyr Pro Phe Leu Leu Met  
355 360 365

Val Val Val Leu Met Ala Ile Leu Ser Phe Gln Val Arg Gln Phe Lys  
370 375 380

Arg Leu Tyr Glu His Ile Lys Asn Asp Lys Tyr Leu Val Gly Gln Arg  
385 390 395 400

Leu Val Asn Tyr Glu Arg Lys Ser Gly Lys Gln Gly Ser Ser Pro Pro  
405 410 415

Pro Pro Gln Ser Ser Gln Glu  
420

30 (2) INFORMATION FOR SEQ ID NO: 400:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 78 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 400:

Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met  
1 5 10 15

Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Leu Thr  
20 25 30

Val Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Cys Leu Ala  
35 40 45

Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro  
50 55 60

Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu Xaa  
65 70 75

55 (2) INFORMATION FOR SEQ ID NO: 401:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 74 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 401:

Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser  
1 5 10 15

His Cys Trp Gly Leu Pro Leu His Val Ala Pro Leu Cys Arg Gly His  
20 25 30

Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala Trp  
35 40 45

Asn Arg Asn Leu Ala Asn Gln Arg His Phe Cys Pro Ser Ile Phe  
50 55 60

His Thr Cys Pro Thr Val Leu Phe Xaa  
65 70

20 (2) INFORMATION FOR SEQ ID NO: 402:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 402:

Ala Arg Thr Ile Leu Val Leu Tyr Leu Ser Leu Gln Arg Leu Glu Asn  
1 5 10 15

Leu Ala Tyr His  
20

35 (2) INFORMATION FOR SEQ ID NO: 403:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 87 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 403:

Met Pro Leu Pro Ser Val Pro Ile Leu Gly Ile Phe Ser Phe Leu Ile  
1 5 10 15

Pro Ser Ser Gln Gly Val Ser Tyr Thr Lys Leu Pro Ile Ser Ser Pro  
20 25 30

Gln Tyr Ser Pro Phe Val Asn Asp His Phe Ser Phe Leu Asn Pro Phe  
35 40 45

Pro Val Gln Ile His Thr Gly Phe Ala Arg Val Gly Ser Tyr Met Gln  
50 55 60

Met Pro Leu Val His Leu Cys Leu Leu Gln Thr Ser Leu Met Lys Asn  
65 70 75 80

Ser Gly Val Gln Gln Gly Ser  
85

- 5 (2) INFORMATION FOR SEQ ID NO: 404:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 92 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 404:  
 Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile Ile  
 1 5 10 15  
 Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala Val  
 20 25 30  
 Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile Thr  
 35 40 45  
 Leu Leu Ile Leu Ala Glu Leu Leu Ile Phe Ser Val Pro Ile Val Tyr  
 50 55 60  
 Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg Asp  
 65 70 75 80  
 Gln Thr Lys Ser Ile Val Glu Lys Ile Pro Ser Lys  
 85 90
- 30 (2) INFORMATION FOR SEQ ID NO: 405:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 21 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 405:  
 Met Ala Cys Ser Cys Leu Met Ile Gln Ser Phe Thr Ser Ala Leu  
 1 5 10 15  
 Val Leu Phe Tyr Gly  
 20
- 45 (2) INFORMATION FOR SEQ ID NO: 406:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 174 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 406:  
 Met Glu Glu Gly Gly Asn Leu Gly Leu Ile Lys Met Val His Leu  
 1 5 10 15  
 Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp Val Thr Phe Val  
 20 25 30

- Ser Gly Phe Pro Ala Phe Pro Lys Pro Ser Pro Thr Tyr Leu Arg Thr  
 35 40 45  
 Ser Ala Glu Gln Thr Leu Pro Leu Leu Pro His Leu His Gly Leu  
 5 50 55 60  
 Cys Leu His Gln Pro Leu His Leu Gly Phe Thr Ala Cys Leu Gly Ser  
 65 70 75 80  
 Ala His Ile Leu Gly Gly Gln Pro Ala Leu Pro Ala Val Pro Glu Pro  
 85 90 95  
 Tyr Ala Gly His Cys Gln Arg Pro Leu Ala Gly Thr Pro His His Ser  
 100 105 110 115  
 Cys His Val Gly Pro Ala Asn Arg Gly Arg Arg Ser Glu Ala Trp Val  
 120 125 130 135 140  
 Gly Arg Tyr Gln Ala Ala Asn Arg Phe Pro Ile Leu Asn Ala Xaa Cys  
 145 150 155 160  
 Thr Met Gly Cys Pro Leu Phe Ala Ile Trp Ala Ala Ser Xaa  
 165 170
- 20 (2) INFORMATION FOR SEQ ID NO: 407:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 64 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 407:  
 Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
 1 5 10 15  
 Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
 20 25 30  
 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
 35 40 45  
 Thr Glu Gly Ile Pro Cys Gln Arg Tyr Gln Asn Gly Leu His Ile Xaa  
 50 55 60
- 55 (2) INFORMATION FOR SEQ ID NO: 408:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 280 amino acids  
 (B) TYPE: amino acid
- 60

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 408:

Met Glu Ala Val Val Asn Leu Tyr Gln Glu Val Met Lys His Ala Asp  
1 5 10 15  
Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser Pro Leu Leu Met Thr  
20 25 30  
Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu Ser Leu Gly Pro Arg  
35 40 45  
Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg Gly Phe Met Ile Val  
50 55 60  
Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr Ile Val Tyr Glu Phe  
65 70 75 80  
Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp Arg Cys Asp Pro Val  
85 90 95  
Asp Tyr Ser Asn Ser Pro Glu Ala Leu Arg Met Val Arg Val Ala Trp  
100 105 110  
Leu Phe Leu Phe Ser Lys Phe Ile Glu Leu Met Asp Thr Val Ile Phe  
115 120 125  
Ile Leu Arg Lys Lys Asp Gly Gln Val Thr Phe Leu His Val Phe His  
130 135 140  
His Ser Val Leu Pro Trp Ser Trp Trp Trp Gly Val Lys Ile Ala Pro  
145 150 155 160  
Gly Gly Met Gly Ser Phe His Ala Met Ile Asn Ser Ser Val His Val  
165 170 175  
Ile Met Tyr Leu Tyr Tyr Gly Leu Ser Ala Phe Gly Pro Val Ala Gln  
180 185 190  
Pro Tyr Leu Trp Trp Lys Lys His Met Thr Ala Ile Gln Leu Ile Gln  
195 200 205  
Phe Val Leu Val Ser Leu His Ile Ser Gln Tyr Tyr Phe Met Ser Ser  
210 215 220  
Cys Asn Tyr Gln Tyr Pro Val Ile Ile His Leu Ile Trp Met Tyr Gly  
225 230 235 240  
Thr Ile Phe Phe Met Leu Phe Ser Asn Phe Trp Tyr His Ser Tyr Thr  
245 250 255  
Lys Gly Lys Arg Leu Pro Arg Ala Leu Gln Gln Asn Gly Ala Pro Gly  
260 265 270  
55 Ile Ala Lys Val Lys Ala Asn Xaa  
275 280  
60 (2) INFORMATION FOR SEQ ID NO: 409:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 284 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 409:

Met Xaa Leu Trp Pro Gln Thr Cys Ser Gly Lys Phe Asp Gly Thr Leu  
1 5 10 15  
Ala Phe Ser Ile His Xaa Leu Ala Val Ile Leu Gly Asp Gln Leu Thr  
20 25 30  
Ala Ala Asp Leu Val Pro Ile Phe Asn Gly Phe Leu Lys Asp Leu Asp  
35 40 45  
Glu Val Arg Ile Gly Val Leu Lys His Leu His Asp Phe Leu Lys Leu  
50 55 60  
Leu His Ile Asp Lys Arg Arg Glu Tyr Leu Tyr Gln Leu Gln Glu Phe  
65 70 75 80  
Leu Val Thr Asp Asn Ser Arg Asn Trp Arg Phe Arg Ala Glu Leu Ala  
85 90 95  
Glu Gln Leu Ile Leu Leu Glu Leu Tyr Ser Pro Arg Asp Val Tyr  
100 105 110  
Asp Tyr Leu Arg Pro Ile Ala Leu Asn Leu Cys Ala Asp Lys Val Ser  
115 120 125  
Ser Val Arg Trp Ile Ser Tyr Lys Leu Val Ser Glu Met Val Lys Lys  
130 135 140  
Leu His Ala Ala Thr Pro Thr Thr Phe Gly Val Asp Leu Ile Asn Glu  
145 150 155 160  
Leu Val Glu Asn Phe Gly Arg Cys Pro Lys Trp Ser Gly Arg Gln Ala  
165 170 175  
Phe Val Phe Val Cys Gln Thr Val Ile Glu Asp Asp Cys Leu Pro Met  
180 185 190  
Asp Gln Phe Ala Val His Leu Met Pro His Leu Leu Thr Leu Ala Asn  
195 200 205  
Asp Arg Val Pro Asn Val Arg Val Leu Leu Ala Lys Thr Leu Arg Gln  
210 215 220  
Thr Leu Leu Glu Lys Asp Tyr Phe Leu Ala Ser Ala Ser Cys His Gln  
225 230 235 240  
Glu Ala Val Glu Gln Thr Ile Met Ala Leu Gln Met Asp Arg Asp Ser  
245 250 255  
Asp Val Lys Tyr Phe Ala Ser Ile His Pro Ala Ser Thr Lys Ile Ser  
260 265 270  
Glu Asp Ala Met Ser Thr Ala Ser Thr Tyr Xaa  
275 280

600

601

(2) INFORMATION FOR SEQ ID NO: 410:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 187 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 410:

Met Leu Phe Leu Phe Phe Val Ile Ile Phe Leu Phe Val Phe Leu Ile 15  
1 5 10  
Leu Ile Ile Gln Phe Ser Lys Pro Leu Thr Asn Pro His Pro Pro Ala 30  
20  
Gly Xaa Ser Asp Arg Arg Arg Tyr Ser Ser Tyr Arg Ser His Asp 45  
35 40  
His Tyr Gln Arg Gln Arg Val Leu Gln Lys Glu Arg Ala Ile Glu Glu 60  
50 55  
Arg Arg Val Val Phe Ile Gly Lys Ile Pro Gly Arg Met Thr Arg Ser 80  
65 70  
Glu Leu Lys Gln Arg Phe Ser Val Phe Gly Glu Ile Glu Glu Cys Thr 95  
85 90  
Ile His Phe Arg Val Gln Gly Asp Asn Tyr Gly Phe Val Thr Tyr Arg 110  
100 105  
Tyr Ala Glu Glu Ala Phe Ala Ile Glu Ser Gly His Lys Leu Arg 125  
115 120  
Gln Ala Asp Glu Gln Pro Phe Asp Leu Cys Phe Gly Gly Arg Arg Xaa 140  
130 135  
Xaa Cys Lys Arg Ser Tyr Ser Asp Leu Asp Ser Asn Arg Glu Asp Phe 155  
145 150  
Asp Pro Ala Pro Val Lys Ser Lys Phe Asp Ser Leu Asp Phe Asp Thr 175  
165 170  
Leu Leu Lys Gln Ala Gln Lys Asn Leu Arg Arg 185  
180

(2) INFORMATION FOR SEQ ID NO: 411:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 237 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 411:

Met Lys Leu Pro Gly Lys Phe Arg Arg Ala His Gln Gly Asn Leu Glu 15  
1 5 10

Ser Gln Leu Thr Ser Glu Ser Tyr Tyr Lys Glu Thr Leu Ser Val Pro 30  
20 25  
Thr Val Glu His Ile Ile Gln Glu Leu Lys Asp Ile Phe Ser Glu Gln 45  
35 40  
His Leu Lys Ala Leu Lys Cys Leu Ser Leu Val Pro Ser Val Met Gly 60  
50 55  
Gln Leu Lys Phe Asn Thr Ser Glu Glu His Ala Asp Met Tyr Arg 80  
65 70  
Ser Asp Leu Pro Asn Pro Asp Thr Leu Ser Ala Glu Leu His Cys Trp 95  
85 90  
Arg Ile Lys Trp Lys His Arg Gly Lys Asp Ile Glu Leu Pro Ser Thr 110  
100 105  
Ile Tyr Glu Ala Leu His Leu Pro Asp Ile Lys Phe Phe Pro Asn Val 125  
115 120  
Tyr Ala Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu 140  
130 135  
Asn Glu Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg 160  
145 150  
Asn Thr Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile 175  
165 170  
Asn Phe Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile 190  
180 185  
Lys Leu Tyr Thr Xaa Xaa Ser Xaa Leu Xaa Thr Xaa Xaa Ser Xaa Xaa 205  
195 200  
Val Glu Xaa Xaa Xaa Xaa Xaa Xaa Xaa Xaa Gly Xaa Xaa Xaa Xaa 220  
210 215  
Asp Xaa Xaa Xaa Arg Glu Lys Ala Val Arg Cys Met Xaa 235  
225 230

(2) INFORMATION FOR SEQ ID NO: 412:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 192 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 412:

Met Lys Pro Met Ala Val Val Ala Ser Thr Val Leu Gly Leu Val Gln 15  
1 5 10  
Asn Met Arg Ala Phe Gly Gly Ile Leu Val Val Val Tyr Tyr Val Phe 30  
20 25  
Ala Ile Ile Gly Ile Asn Leu Phe Arg Gly Val Ile Val Ala Leu Pro 45  
35 40

Gly Asn Ser Ser Leu Ala Pro Ala Asn Gly Ser Ala Pro Cys Gly Ser  
50 55 60

5 Phe Glu Gln Leu Glu Tyr Trp Ala Asn Asn Phe Asp Asp Phe Ala Ala  
65 70 75 80

Ala Leu Val Thr Leu Trp Asn Leu Met Val Val Asn Asn Trp Gln Val  
85 90 95

10 Phe Leu Asp Ala Tyr Arg Arg Tyr Ser Gly Pro Trp Ser Lys Ile Tyr  
100 105 110

15 Phe Val Leu Trp Trp Leu Val Ser Ser Val Ile Trp Val Asn Leu Phe  
115 120 125

Leu Ala Leu Ile Leu Glu Asn Phe Leu His Lys Lys Trp Asp Pro Arg Ser  
130 135 140

20 His Leu Gln Pro Leu Ala Gly Thr Pro Glu Ala Thr Tyr Gln Met Thr  
145 150 155 160

Val Glu Leu Leu Phe Arg Asp Ile Leu Glu Glu Pro Gly Glu Asp Glu  
165 170 175

25 Leu Thr Glu Arg Leu Ser Gln His Pro His Leu Trp Leu Cys Arg Xaa  
180 185 190

30

(2) INFORMATION FOR SEQ ID NO: 413:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 413:

Asn Val Val Val Val Ala Phe Gly Leu Ile Leu Ile Ile Glu Ser Leu  
1 5 10 15

45

Gly Glu Gln Cys Pro  
20

(2) INFORMATION FOR SEQ ID NO: 414:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 414:

Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Tyr Glu Lys Lys Lys  
1 5 10 15

60

Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30

5 Ser Asp Gly Ile Leu Gln Phe Glu Ser Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45

Ala Pro Xaa  
50

10

(2) INFORMATION FOR SEQ ID NO: 415:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 415:

20 Met Leu Ile Ile Ser Leu Arg Pro Gln Phe Pro Ser Leu Ile Val Gln  
1 5 10 15

Leu Glu Cys Ser Val Leu Phe Leu Pro Ile Ser Leu Asn Leu Leu Leu  
20 25 30

25

30

(2) INFORMATION FOR SEQ ID NO: 416:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 163 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 416:

Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15

Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30

45

Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45

Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60

50

Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75 80

55

Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95

60

Glu Ile Arg Gly Gly Glu Leu Lys Tyr Cys Asn Leu Leu Val Arg  
100 105 110

60

Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

115 120 125  
Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Arg  
130 135 140  
5 Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Ile Gly Arg Pro  
145 150 155 160  
Gln Val Xaa  
10

(2) INFORMATION FOR SEQ ID NO: 417:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 174 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 417:  
Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15  
25 Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30  
Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Tyr Ser Ser  
35 40 45  
30 Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60  
Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Phe Ser  
65 70 75 80  
Glu Asp Gly Ala Leu Met Asp Gly Gly Met Asp Leu Asn Met Glu Gln  
85 90 95  
40 Gly Met Ala Glu His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val  
100 105 110  
Gln Val Leu Ser Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu  
115 120 125  
45 Ala Pro Gly Arg Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro  
130 135 140  
Trp Phe Thr Ala Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys  
145 150 155 160  
Arg Gln Arg Arg Gln Glu Arg Arg Gln Met Lys Arg Leu Xaa  
165 170  
50  
55

(2) INFORMATION FOR SEQ ID NO: 418:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 50 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 418:  
5 Met Glu Leu Pro Lys Gly Leu Gln Gly Val Gly Pro Val Ala Met Met  
1 5 10 15  
Arg Pro Phe Tyr Leu Leu Leu Pro Val Leu Cys Thr Gln Ala Leu Arg  
20 25 30  
10 Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Cys Cys Leu  
35 40 45  
Ala Xaa  
15 50

(2) INFORMATION FOR SEQ ID NO: 419:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 120 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 419:  
Met Leu Gly Lys Gly Gly Arg Ala Gly Leu Leu Arg Tyr Arg Leu  
1 5 10 15  
30 Leu Tyr Phe Thr Leu Val Val Gly Glu Gly Glu Pro Gly Glu Asn Lys  
20 25 30  
Val Thr Ile Pro Phe Phe Glu Thr Gly Lys Lys Ile Ile Phe Cys Ser  
35 40 45  
35 Val Lys Met Val Glu Asn Ser Asn Val Pro Ser His Lys Gly Pro Val  
50 55 60  
Pro Leu Arg Ser Glu Gln Trp Glu Leu Lys Ile Ser Glu Thr Leu Gly  
65 70 75 80  
Glu Gly Lys Ile Gly Phe Leu Leu Ile Gly Arg Cys Ser Ser Gly Xaa  
85 90 95  
45 Gly Gly Leu Cys Phe Cys Trp Asp Val Leu Cys Cys Met Tyr Ala Tyr  
100 105 110  
Met Asp Arg Ser Leu Leu Ser Leu  
115 120  
50

(2) INFORMATION FOR SEQ ID NO: 420:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 159 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 420:  
55  
60

Met Thr His Leu Leu Leu Thr Ala Thr Val Thr Pro Ser Glu Gln Asn  
1 5 10 15  
Ser Ser Arg Glu Pro Gly Trp Glu Thr Ala Met Ala Lys Asp Ile Leu  
20 25 30  
Gly Glu Ala Gly Leu His Phe Asp Glu Leu Asn Lys Leu Arg Val Leu  
35 40 45  
Asp Pro Glu Val Thr Gln Gln Thr Ile Glu Leu Lys Glu Gln Cys Lys  
50 55 60  
Asp Phe Val Asp Lys Ile Gly Gln Phe Gln Lys Ile Val Gly Gly Leu  
65 70 75  
Ile Glu Leu Val Asp Gln Leu Ala Lys Glu Ala Glu Asn Glu Lys Met  
80 85 90 95  
Lys Ala Ile Gly Ala Arg Asn Leu Leu Lys Ser Ile Ala Lys Gln Arg  
100 105 110  
Glu Ala Gln Gln Gln Leu Gln Ala Leu Ile Ala Glu Lys Lys Met  
115 120 125  
Gln Leu Glu Arg Tyr Arg Val Glu Tyr Glu Ala Leu Cys Lys Val Glu  
130 135 140  
Ala Glu Gln Asn Glu Phe Ile Asp Gln Phe Ile Phe Gln Lys Xaa  
145 150 155  
(2) INFORMATION FOR SEQ ID NO: 421:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 154 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 421:  
Met Asn Val Gly Val Ala His Ser Glu Val Asn Pro Asn Thr Arg Val  
1 5 10 15  
Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu  
20 25 30  
Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala  
35 40 45  
Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu  
50 55 60  
His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala  
65 70 75  
Arg Leu Leu Thr His Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr  
80 85 90 95  
Ser Ser Arg Lys Phe Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu  
100 105 110

Ala Ser Phe Tyr Thr Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr  
115 120 125  
5 Ala Ser Leu Leu Ser Val Leu Ile Pro Lys Met Pro Gln Leu His Gly  
130 135 140  
Val Arg Ile Phe Gly Ile Asn Lys Tyr Xaa  
145 150  
10  
(2) INFORMATION FOR SEQ ID NO: 422:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 204 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 422:  
Met Val Cys Gly Gly Phe Ala Cys Ser Lys Asn Cys Leu Cys Ala Leu  
1 5 10 15  
Asn Leu Leu Tyr Thr Leu Val Ser Leu Leu Ile Gly Ile Ala Ala  
20 25 30  
Trp Gly Ile Gly Phe Gly Leu Ile Ser Ser Leu Arg Val Val Gly Val  
35 40 45  
Val Ile Ala Val Gly Ile Phe Leu Phe Leu Ile Ala Leu Val Gly Leu  
50 55 60  
Ile Gly Ala Val Lys His Gln Val Leu Leu Phe Phe Tyr Met Ile  
65 70 75 80  
Ile Leu Leu Val Phe Ile Val Gln Phe Ser Val Ser Cys Ala Cys  
85 90 95  
Leu Ala Leu Asn Gln Glu Gln Gln Gly Gln Leu Leu Glu Val Gly Trp  
100 105 110  
Asn Asn Thr Ala Ser Ala Arg Asn Asp Ile Gln Arg Asn Leu Asn Cys  
115 120 125  
Cys Gly Phe Arg Ser Val Asn Pro Asn Asp Thr Cys Leu Ala Ser Cys  
130 135 140  
Val Lys Ser Asp His Ser Cys Ser Pro Cys Ala Pro Ile Ile Gly Glu  
145 150 155 160  
Tyr Ala Gly Glu Val Leu Arg Phe Val Gly Ile Gly Leu Phe Phe  
165 170 175  
Ser Phe Thr Glu Ile Leu Gly Val Trp Leu Thr Tyr Arg Tyr Arg Asn  
180 185 190  
Gln Lys Asp Pro Arg Ala Asn Pro Ser Ala Phe Leu  
195 200  
60



## (2) INFORMATION FOR SEQ ID NO: 423:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 67 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 423:

10 Met Leu Gln Ser Ile Ile Lys Asn Ile Trp Ile Pro Met Lys Pro Tyr 15  
 1 Tyr Thr Lys Val Tyr Gln Glu Ile Trp Ile Gly Met Gly Leu Met Gly 30  
 15 Phe Ile Val Tyr Lys Ile Arg Ala Ala Asp Lys Arg Ser Lys Ala Leu 45  
 20 Lys Ala Ser Ala Pro Ala Pro Gly His His Asn Gln Ile Tyr Leu Glu 60  
 Tyr Met Xaa 65

## (2) INFORMATION FOR SEQ ID NO: 424:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 25 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 424:

35 Met Leu Gly Val Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val 15  
 40 Ala Val Asn Asn Pro Lys Lys Gln Glu 25

## (2) INFORMATION FOR SEQ ID NO: 425:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 299 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 425:

45 Met Ala Ala Xaa Glu Pro Ala Val Leu Ala Leu Pro Asn Ser Gly Ala 15  
 50 Gly Gly Ala Gly Ala Pro Ser Gly Thr Val Pro Val Leu Phe Cys Phe 30  
 55 Ser Val Phe Ala Arg Pro Ser Val Pro His Gly Ala Gly Tyr Glu 45  
 60 Leu Leu Ile Gln Lys Phe Leu Ser Leu Tyr Gly Asp Gln Ile Asp Met 10

50 55 60  
 His Arg Lys Phe Val Val Gln Leu Phe Ala Glu Glu Trp Gly Gln Tyr 65  
 5 Val Asp Leu Pro Lys Gly Phe Ala Val Ser Glu Arg Cys Lys Val Arg 85  
 10 Leu Val Pro Leu Gln Ile Gln Leu Thr Thr Leu Gly Asn Leu Thr Pro 100  
 110 Ser Ser Thr Val Phe Phe Cys Asp Met Gln Glu Arg Phe Arg Pro 115  
 120 125  
 15 Ala Ile Lys Tyr Phe Gly Asp Ile Ile Ser Val Gly Gln Arg Leu 130  
 135 140  
 20 Gln Gly Ala Arg Ile Leu Gly Ile Pro Val Ile Val Thr Glu Gln Tyr 145  
 150 155 160  
 Pro Lys Gly Leu Gly Ser Thr Val Gln Glu Ile Asp Leu Thr Gly Val 165  
 170 175  
 25 Lys Leu Val Leu Pro Lys Thr Lys Phe Ser Met Val Leu Pro Glu Val 180  
 185 190  
 Glu Ala Ala Leu Ala Glu Ile Pro Gly Val Arg Ser Val Val Leu Phe 195  
 200 205  
 30 Gly Val Glu Thr His Val Cys Ile Gln Gln Thr Ala Leu Glu Leu Val 210  
 215 220  
 Gly Arg Gly Val Glu Val His Ile Val Ala Asp Ala Thr Ser Ser Arg 225  
 230 235 240  
 Ser Met Met Asp Arg Met Phe Ala Leu Glu Arg Leu Ala Xaa Xaa Gly 245  
 250 255  
 40 Ile Ile Val Thr Thr Ser Glu Ala Val Leu Leu Gln Leu Val Ala Asp 260  
 265 270  
 Lys Asp His Pro Lys Phe Lys Glu Ile Gln Asn Leu Ile Lys Ala Ser 275  
 280 285  
 45 Ala Pro Glu Ser Gly Leu Leu Ser Lys Val Xaa 290  
 295  
 50 (2) INFORMATION FOR SEQ ID NO: 426:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 13 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 426:  
 55 Met Arg Asp Leu Gly Thr Leu Leu Ser Pro Val Cys Ser 10  
 1 5 10

## (2) INFORMATION FOR SEQ ID NO: 427:

5

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 198 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 427:

10 Met Phe Gly Cys Leu Val Ala Gly Arg Leu Val Gln Thr Ala Ala Gln  
 1 5 10 15

15 Gln Val Ala Glu Asp Lys Phe Val Phe Asp Leu Pro Asp Tyr Glu Ser  
 20 25 30

Ile Asn His Val Val Val Phe Met Leu Gly Thr Ile Pro Phe Pro Glu  
 35 40 45

20 Gly Met Gly Gly Ser Val Tyr Phe Ser Tyr Pro Asp Ser Asn Gly Met  
 50 55 60

Pro Val Trp Gln Leu Leu Gly Phe Val Thr Asn Gly Lys Pro Ser Ala  
 65 70 75 80

25 Ile Phe Lys Ile Ser Gly Leu Lys Ser Gly Glu Gly Ser Gln His Pro  
 85 90 95

30 Phe Gly Ala Met Asn Ile Val Arg Thr Pro Ser Val Ala Gln Ile Gly  
 100 105 110

Ile Ser Val Glu Leu Asp Ser Met Ala Gln Gln Thr Pro Val Gly  
 115 120 125

35 Asn Ala Ala Val Ser Val Asp Ser Phe Thr Gln Phe Thr Gln Lys  
 130 135 140

Met Leu Asp Asn Phe Tyr Asn Phe Ala Ser Ser Phe Ala Val Ser Gln  
 145 150 155 160

40 Ala Gln Met Thr Pro Ser Pro Ser Glu Met Phe Ile Pro Ala Asn Val  
 165 170 175

45 Val Leu Lys Trp Tyr Glu Asn Phe Gln Arg Arg Leu Ala Gln Asn Pro  
 180 185 190

Xaa Phe Trp Xaa Thr Xaa  
 195

50

## (2) INFORMATION FOR SEQ ID NO: 428:

55

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 47 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 428:

60 Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser

1

5

10

15

Ala Gly Val Asn Phe Ile Leu Pro Leu Leu Leu Leu Trp Lys  
 20 25 30

5

Asn Arg Gly Gly Val Gly Arg Ser Val Met Ser Ala Val Glu Xaa  
 35 40 45

10

## (2) INFORMATION FOR SEQ ID NO: 429:

15

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 370 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 429:

20 Met Lys Lys Val Glu Glu Lys Arg Val Asp Val Asn Ser Ala Val Ala  
 1 5 10 15

Met Gly Glu Val Ile Leu Ala Val Cys His Pro Asp Cys Ile Thr Thr  
 20 25 30

25 Ile Lys His Trp Ile Thr Ile Ile Arg Ala Arg Phe Glu Glu Val Leu  
 35 40 45

Thr Trp Ala Lys Gln His Gln Arg Leu Glu Thr Ala Leu Ser Glu  
 50 55 60

30 Leu Val Ala Asn Ala Glu Leu Leu Glu Glu Leu Ala Trp Ile Gln  
 65 70 75 80

35 Trp Ala Glu Thr Thr Leu Ile Gln Arg Asp Gln Glu Pro Ile Pro Gln  
 85 90 95

Asn Ile Asp Arg Val Lys Ala Leu Ile Ala Glu His Gln Thr Phe Met  
 100 105 110

40 Glu Glu Met Thr Arg Lys Gln Pro Asp Val Asp Arg Val Thr Lys Thr  
 115 120 125

Tyr Lys Arg Lys Asn Ile Glu Pro Thr His Ala Pro Phe Ile Glu Lys  
 130 135 140

Ser Arg Ser Gly Gly Arg Lys Ser Leu Ser Gln Pro Thr Pro Pro  
 145 150 155 160

50 Met Pro Ile Leu Ser Gln Ser Glu Ala Lys Asn Pro Arg Ile Asn Gln  
 165 170 175

Leu Ser Ala Arg Trp Gln Gln Val Trp Leu Leu Ala Leu Glu Arg Gln  
 180 185 190

55 Arg Lys Leu Asn Asp Ala Leu Asp Arg Leu Glu Leu Lys Glu Phe  
 195 200 205

Ala Asn Phe Asp Phe Asp Val Trp Arg Lys Tyr Met Arg Trp Met  
 210 215 220

60

Asn His Lys Lys Ser Arg Val Met Asp Phe Phe Arg Arg Ile Asp Lys  
225 230 235 240

Asp Gln Asp Gly Lys Ile Thr Arg Gln Glu Phe Ile Asp Gly Ile Leu  
245 250 255

Ala Ser Lys Phe Pro Thr Thr Lys Leu Glu Met Thr Ala Val Ala Asp  
260 265 270

10 Ile Phe Asp Arg Asp Gly Tyr Ile Asp Tyr Tyr Glu Phe Val  
275 280 285

Ala Ala Leu His Pro Asn Lys Asp Ala Tyr Arg Pro Thr Thr Asp Ala  
290 295 300

15 Asp Lys Ile Glu Asp Glu Val Thr Arg Gln Val Ala Gln Cys Lys Cys  
305 310 315 320

Ala Lys Arg Phe Gln Val Glu Gln Ile Gly Glu Asn Lys Tyr Arg Phe  
325 330 335

Phe Leu Gly Asn Gln Phe Gly Asp Ser Gln Gln Leu Arg Leu Val Arg  
340 345 350

25 Ile Leu Arg Asn Arg Asp Gly Ser Arg Trp Trp Arg Met Asp Gly Leu  
355 360 365

Gly Xaa  
370

(2) INFORMATION FOR SEQ ID NO: 430:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 430:

40 Met Asn Val Lys Thr Phe Ser Xaa Asp His Met His Phe Leu Cys Cys  
1 5 10 15

45 Leu Tyr Leu Arg Tyr Val Thr Phe Val Tyr Leu Asn Leu Phe  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 431:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 431:

55 Met Glu Pro His Leu Arg Cys Arg Val Thr Arg Val Arg Gly Ser Leu  
1 5 10 15

60 Gly Asn Thr Gly Arg Trp Leu Leu

20

5 (2) INFORMATION FOR SEQ ID NO: 432:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 53 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 432:

15 Met His Tyr Leu Val Leu Gly Gly Leu Gly Val Phe Leu Phe Phe Ser  
1 5 10 15

Cys Phe Val Phe Leu Phe Phe Xaa Phe Ser Phe Ala Phe Phe Pro Phe  
20 25 30

20 Tyr Leu Glu Gly Met Gly Gly Ser Gly Asn Arg Glu Val Gly Gly Gly  
35 40 45

Phe Cys Leu Phe Phe  
50

(2) INFORMATION FOR SEQ ID NO: 433:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 176 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 433:

35 Met Val Ser Lys Ala Leu Leu Arg Leu Val Ser Ala Val Asn Arg Arg  
1 5 10 15

Arg Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser  
20 25 30

40 Val Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu  
35 40 45

45 Leu Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu  
50 55 60

Lys Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val  
65 70 75 80

50 Lys Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Thr Gly Gly  
85 90 95

Ala Gly Phe Val Gly Ser His Leu Thr Asp Lys Leu Met Met Asp Gly  
100 105 110

55 His Glu Val Thr Val Val Asp Asn Phe Phe Thr Gly Arg Lys Arg Asn  
115 120 125

60 Val Glu His Trp Ile Gly His Glu Asn Phe Glu Leu Ile Asn His Asp  
130 135 140

Val Trp Ser Pro Ser Thr Arg Leu Thr Arg Tyr Thr Ile Trp His  
145 150 155

5 Leu Gln Pro Pro Leu Gln Thr Thr Cys Ile Ile Leu Ser Arg His Xaa  
165 170 175

10

## (2) INFORMATION FOR SEQ ID NO: 434:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 434:

20 Met Leu Arg Cys Trp Pro Leu Phe Trp Leu Pro Leu Val Ser Pro Phe  
1 5 10 15

25 Cys Ser Leu Phe Trp Leu Val Glu Trp Phe Gly Thr Asn Ile Asp  
20 25 30

Arg Glu Ser Tyr Asp Ala Ile Gly Gly Pro Ser Trp Met Thr Ala Ser  
35 40 45

30 Ser Phe Cys Leu Ser Asn Ser Asn Ile Trp Ser Leu Glu Ile Ser Ser  
50 55 60

Gly Ser Thr Val Val His Ser Gln Gln Ala Met Asp  
65 70 75

35

## (2) INFORMATION FOR SEQ ID NO: 435:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 32 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 435:

45 Met Arg Ser Cys Glu Ile Gln Leu Cys Val Trp Leu Leu Val Ser Ser  
1 5 10 15

50 His Val Asp Met Val Leu Gly Gly Ser Pro Ser Thr Leu Tyr Met Met  
20 25 30

55

## (2) INFORMATION FOR SEQ ID NO: 436:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 436:

5 Met Val Val Asn Ser Leu Cys Phe Leu Ser Leu Leu Val Ile Leu  
1 5 10 15

10 Glu Leu Ser Thr Asp Ser Ser Ala Arg Leu Leu Tyr His Glu  
20 25 30

## (2) INFORMATION FOR SEQ ID NO: 437:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 437:

20 Met Asp Lys Gln Lys His Leu Glu Val Arg Arg Ser Val Phe Lys Ile  
1 5 10 15

25 Gln Gly Lys Ile Ala Phe Ser Leu Met Phe Val Leu Lys Asp Leu Ser  
20 25 30

Pro Thr Ile Phe Ser His Ser Ile Leu Leu Leu Pro His His Val  
35 40 45

30 Leu Pro Cys Thr Pro Gln Met Val Arg Gly Val Thr Gln Val Leu Arg  
50 55 60

Glu Phe Gly Asp Gln  
65

35

## (2) INFORMATION FOR SEQ ID NO: 438:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 438:

45 Met Pro Leu Cys Phe Phe Ser Phe Leu Cys Cys Trp Val Leu Val Phe  
1 5 10 15

Lys Leu Ile  
50

## (2) INFORMATION FOR SEQ ID NO: 439:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 43 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 439:

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15

5 Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg  
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe Xaa  
35 40

10

(2) INFORMATION FOR SEQ ID NO: 440:

15 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 440:

20 Met Leu Leu Phe Pro Ser Leu Leu Phe Ala Ala Thr Tyr Asn Val Ala  
1 5 10 15

25 Asn Pro Ser Arg Leu Ile Leu Tyr Met Ile Ser Ala Gly Ala Asp Ser  
20 25 30

Gln

(2) INFORMATION FOR SEQ ID NO: 441:

35 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 53 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 441:

40 Met Trp Gln Val Arg Gly Leu Pro Pro Val Pro Leu Leu Thr Met  
1 5 10 15

Ser Pro Pro Cys Leu Ser Ser Pro Phe Pro Phe Ile Ser Val Pro  
20 25 30

45 Leu Phe Glu Ala Val Pro Ile Ser Val Ser Asp Gln Pro Ser Pro Xaa  
35 40 45

50 Leu Thr Thr Leu Leu  
50

(2) INFORMATION FOR SEQ ID NO: 442:

55 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 442:

60

Met Ile Thr Ser Val Leu Val Phe Leu Ile Phe Phe Pro Tyr Leu  
1 5 10 15

5 Ser Leu Val Thr Leu Leu Gln Ala Arg Asn Leu Trp Val Ile His Arg  
20 25 30

Ala Ala Leu Cys Glu Ser Gly Leu Phe His Trp Arg Lys Gly Ile Glu  
35 40 45

10

Asn Gln Leu Glu Pro Met Tyr Phe Leu Pro His Gly Thr Leu Phe Leu  
50 55 60

15

(2) INFORMATION FOR SEQ ID NO: 443:

20 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 443:

Met Leu Tyr Ser Cys Glu Pro Tyr Leu Ile Ile Leu Asn Ile Tyr Ser  
1 5 10 15

30 Gln Lys Ala Phe Tyr Phe Phe Glu Gly Ser Phe Ser Val Cys  
20 25 30

Thr Leu

35

(2) INFORMATION FOR SEQ ID NO: 444:

40 (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 89 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 444:

Met Arg Gln Arg Gln Ala Ala Cys Gln Pro Pro Ser Arg Asn Gly  
1 5 10 15

50 Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val  
20 25 30

Lys Leu Leu Phe Ile Pro Trp Trp Ala Ser Leu Leu Ser Ser Pro Leu  
35 40 45

55 Asn Leu Leu Leu Val Ser Ile Ser Trp Asp Leu Gly Leu Lys Leu  
50 55 60

60 Asn Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr  
65 70 75 80

Lys Lys Phe Asn Lys Lys Lys Lys Lys  
85

5

(2) INFORMATION FOR SEQ ID NO: 445:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 350 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 445:

Met Asp Phe Ile Thr Ser Thr Ala Ile Leu Pro Leu Leu Phe Gly Cys  
1 5 10 15

Leu Gly Val Phe Gly Leu Phe Arg Leu Leu Gln Trp Val Arg Gly Lys  
20 25 30

Ala Tyr Leu Arg Asn Ala Val Val Ile Thr Gly Ala Thr Ser Gly  
35 40 45

Leu Gly Lys Glu Cys Ala Lys Val Phe Tyr Ala Ala Gly Ala Lys Leu  
50 55 60

Val Leu Cys Gly Arg Asn Gly Gly Ala Leu Glu Glu Leu Ile Arg Glu  
65 70 75 80

Leu Thr Ala Ser His Ala Thr Lys Val Gln Thr His Lys Pro Tyr Leu  
85 90 95

Val Thr Phe Asp Leu Thr Asp Ser Gly Ala Ile Val Ala Ala Ala  
100 105 110

Glu Ile Leu Gln Cys Phe Gly Tyr Val Asp Ile Leu Val Asn Ala  
115 120 125

Gly Ile Ser Tyr Arg Gly Thr Ile Met Asp Thr Thr Val Asp Val Asp  
130 135 140

Lys Arg Val Met Glu Thr Asn Tyr Phe Gly Pro Val Ala Leu Thr Lys  
145 150 155 160

Ala Leu Leu Pro Ser Met Ile Lys Arg Arg Gln Gly His Ile Val Ala  
165 170 175

Ile Ser Ser Ile Gln Gly Lys Met Ser Ile Pro Phe Arg Ser Ala Tyr  
180 185 190

Ala Ala Ser Lys His Ala Thr Gln Ala Phe Phe Asp Cys Leu Arg Ala  
195 200 205

Glu Met Glu Gln Tyr Glu Ile Glu Val Thr Val Ile Ser Pro Gly Tyr  
210 215 220

Ile His Thr Asn Leu Ser Val Asn Ala Ile Thr Ala Asp Gly Ser Arg  
225 230 235 240

Tyr Gly Val Met Asp Thr Thr Thr Ala Gln Gly Arg Ser Pro Val Glu  
245 250 255

60

Val Ala Gln Asp Val Leu Ala Ala Val Gly Lys Lys Lys Lys Asp Val  
260 265 270

Ile Leu Ala Asp Leu Leu Pro Ser Leu Ala Val Tyr Leu Arg Thr Leu  
275 280 285

Ala Pro Gly Leu Phe Phe Ser Leu Met Pro Pro Gly Pro Glu Lys Ser  
290 295 300

Gly Asn Pro Arg Thr Pro Ser Thr Leu Thr Ser Gln Gly Gln Gly Arg  
305 310 315 320

Glu Ala Ala Leu Leu Gly Leu Thr Leu Gln Gly Thr Val Ala Phe  
325 330 335

Val Glu Thr Leu Met Glu Ile Cys Leu Thr Ser Gly Lys Asp  
340 345 350

20

(2) INFORMATION FOR SEQ ID NO: 446:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 49 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 446:

Met Val Phe Leu Pro Arg Gly Val Val Ser Gly Gly Ala Ala Cys  
1 5 10 15

Leu Trp Leu Thr Phe Ile Leu Glu Thr Glu Val Tyr Leu Asp Leu Ala  
20 25 30

Thr Glu Ala Arg Ala His Ser Arg Met Gly Leu Gly Leu Trp Pro Pro  
35 40 45

Asn

40

(2) INFORMATION FOR SEQ ID NO: 447:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 278 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 447:

Met Ala Ser Ala Glu Leu Asp Tyr Thr Ile Glu Ile Pro Asp Gln Pro  
1 5 10 15

Cys Trp Ser Gln Lys Asn Ser Pro Ser Pro Gly Gly Lys Glu Ala Glu  
20 25 30

Thr Arg Gln Pro Val Val Ile Leu Leu Gly Trp Gly Gly Cys Lys Asp  
35 40 45

60

Lys Asn Leu Ala Lys Tyr Ser Ala Ile Tyr His Lys Arg Gly Cys Ile  
50 55 60

Val Ile Arg Tyr Thr Ala Pro Trp His Met Val Phe Phe Ser Glu Ser  
65 70 75 80

Leu Gly Ile Pro Ser Leu Arg Val Leu Ala Gln Lys Leu Leu Glu Leu  
85 90 95

10 Leu Phe Asp Tyr Glu Ile Glu Lys Glu Pro Leu Leu Phe His Val Phe  
100 105 110

Ser Asn Gly Gly Val Met Leu Tyr Arg Tyr Val Leu Glu Leu Leu Gln  
115 120 125

15 Thr Arg Arg Phe Cys Arg Leu Arg Val Val Gly Thr Ile Phe Asp Ser  
130 135 140

20 Ala Pro Gly Asp Ser Asn Leu Val Gly Ala Leu Arg Ala Leu Ala Ala  
145 150 155 160

Ile Leu Glu Arg Arg Ala Ala Met Leu Arg Leu Leu Leu Val Ala  
165 170 175

25 Phe Ala Leu Val Val Leu Phe His Val Leu Leu Ala Pro Ile Thr  
180 185 190

Ala Xaa Phe His Thr Phe Tyr Asp Arg Leu Gln Asp Ala Gly Ser  
195 200 205

30 Arg Trp Pro Glu Leu Tyr Tyr Ser Arg Ala Asp Glu Val Val Leu  
210 215 220

35 Ala Arg Asp Ile Glu Arg Met Val Glu Ala Arg Leu Ala Arg Arg Val  
225 230 235 240

Leu Ala Arg Ser Val Asp Phe Val Ser Ser Ala His Val Ser His Leu  
245 250 255

40 Arg Asp Tyr Pro Thr Tyr Thr Ser Leu Cys Val Asp Phe Met Arg  
260 265 270

Asn Cys Val Arg Cys Xaa  
275

(2) INFORMATION FOR SEQ ID NO: 448:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 199 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 448:

Met Ser Phe Ile Phe Asp Trp Ile Tyr Ser Gly Phe Ser Ser Val Leu  
1 5 10 15

Gln Phe Leu Gly Leu Tyr Lys Lys Thr Gly Lys Leu Val Phe Leu Gly  
20 25 30

Leu Asp Asn Ala Gly Lys Thr Thr Leu Leu His Met Leu Lys Asp Asp  
35 40 45

5 Arg Leu Gly Gln His Val Pro Thr Leu His Pro Thr Ser Glu Glu Leu  
50 55 60

10 Thr Ile Ala Gly Met Thr Phe Thr Thr Phe Asp Leu Gly His Val  
65 70 75 80

Gln Ala Arg Arg Val Trp Lys Asn Tyr Leu Pro Ala Ile Asn Gly Ile  
85 90 95

15 Val Phe Leu Val Asp Cys Ala Asp His Glu Arg Leu Leu Glu Ser Lys  
100 105 110

Glu Glu Leu Asp Ser Leu Met Thr Asp Glu Thr Ile Ala Asn Val Pro  
115 120 125

20 Ile Leu Ile Leu Gly Asn Lys Ile Asp Arg Pro Glu Ala Ile Ser Glu  
130 135 140

Glu Arg Leu Arg Glu Met Phe Gly Leu Tyr Gly Gln Thr Thr Gly Lys  
145 150 155 160

25 Gly Ser Ile Ser Leu Lys Glu Leu Asn Ala Arg Pro Leu Glu Val Phe  
165 170 175

30 Met Cys Ser Val Leu Lys Arg Gln Gly Tyr Gly Glu Gly Phe Arg Trp  
180 185 190

Met Ala Gln Tyr Ile Asp Xaa  
195

(2) INFORMATION FOR SEQ ID NO: 449:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 258 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 449:

45 Met Thr Leu Ser Arg Phe Ala Tyr Asn Gly Lys Arg Cys Pro Ser Ser  
1 5 10 15

Tyr Asn Ile Leu Asp Asn Ser Lys Ile Ile Ser Glu Glu Cys Arg Lys  
20 25 30

50 Glu Leu Thr Ala Leu Leu His His Tyr Tyr Pro Ile Glu Ile Asp Pro  
35 40 45

55 His Arg Thr Val Lys Glu Lys Leu Pro His Met Val Glu Trp Thr  
50 55 60

Lys Ala His Asn Leu Cys Gln Gln Lys Ile Gln Lys Phe Gln Ile  
65 70 75 80

60 Ala Gln Val Val Arg Glu Ser Asn Ala Met Leu Arg Glu Gly Tyr Lys  
85

5 Thr Phe Phe Asn Thr Leu Tyr His Asn Asn Ile Pro Leu Phe Ile Phe 85 90 95  
100 105 110  
Ser Ala Gly Ile Gly Asp Ile Leu Glu Ile Ile Arg Gln Met Lys 115 120 125  
130 135 140  
10 Val Phe His Pro Asn Ile His Ile Val Ser Asn Tyr Met Asp Phe Asn 145 150 155 160  
Glu Asp Gly Phe Leu Gln Gly Phe Lys Gly Gln Leu Ile His Thr Tyr 165 170 175  
15 Asn Lys Asn Ser Val Cys Glu Asn Xaa Gly Tyr Phe Gln Gln Leu 180 185 190 195  
Glu Gly Lys Thr Asn Val Ile Leu Leu Gly Asp Ser Ile Gly Asp Leu 200 205 210 215  
20 Thr Met Ala Asp Gly Val Pro Gly Val Gln Asn Ile Leu Lys Ile Gly 220 225 230 235  
Phe Leu Asn Asp Lys Val Glu Arg Xaa Arg Tyr Met Asp Ser 240 245 250 255  
25 Tyr Asp Ile Val Leu Glu Lys Asp Glu Thr Leu Asp Val Val Asn Gly 260 265 270 275  
30 Leu Leu Gln His Ile Leu Cys Gln Gly Val Gln Leu Glu Met Gln Gly 280 285 290 295  
Pro Xaa 300 305 310 315  
35 (2) INFORMATION FOR SEQ ID NO: 450:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 87 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 450:  
Met Ser His Val Leu Leu Cys Pro Ser Leu Ser Cys Ser Asn Leu Leu 1 5 10 15  
Pro Pro Ser His Ser Leu Gly Thr Met Gly Ser Leu Ser Pro His Leu 20 25 30  
Cys Gly His Thr Met Cys Pro Val Asn Pro Glu Leu Pro Leu Ser Ser 35 40 45  
Arg Leu Thr Thr Asp Gln Pro Gln Pro Asp Ala Cys Ser Pro Thr Leu 50 55 60  
Leu Thr Leu Pro Leu Pro Ser Ser Phe Leu Pro His Ser Lys Pro Thr 65 70 75 80

Phe Xaa His Pro Cys Ser Pro 85  
5 (2) INFORMATION FOR SEQ ID NO: 451:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 315 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 451:  
Met Phe Ser Ile Asn Pro Leu Glu Asn Leu Lys Val Tyr Ile Ser Ser 1 5 10 15  
Arg Pro Leu Val Val Phe Met Ile Ser Val Xaa Pro Met Ala Ile 20 25 30  
Ala Phe Leu Thr Leu Gly Tyr Phe Phe Lys Ile Lys Glu Ile Lys Ser 35 40 45  
Pro Glu Met Ala Glu Asp Trp Asn Thr Phe Leu Leu Arg Phe Asn Asp 50 55 60  
Leu Asp Leu Cys Val Ser Glu Asn Glu Thr Leu Lys His Leu Thr Asn 65 70 75 80  
Asp Thr Thr Thr Pro Glu Ser Thr Met Thr Ser Gly Gln Ala Arg Ala 85 90 95  
Ser Thr Gln Ser Pro Gln Ala Leu Glu Asp Ser Gly Pro Val Asn Ile 100 105 110 115  
Ser Val Ser Ile Thr Leu Thr Leu Asp Pro Leu Lys Pro Phe Gly Gly 120 125 130 135  
Tyr Ser Arg Asn Val Thr His Leu Tyr Ser Thr Ile Leu Gly His Gln 140 145 150 155  
Ile Gly Leu Ser Gly Arg Glu Ala His Glu Glu Ile Asn Ile Thr Phe 160 165 170 175  
Thr Leu Pro Thr Ala Trp Ser Ser Asp Asp Cys Ala Leu His Gly His 180 185 190 195  
Cys Glu Gln Val Val Phe Thr Ala Cys Met Thr Leu Thr Ala Ser Pro 200 205 210 215  
Gly Val Phe Pro Val Thr Val Gln Pro Pro His Cys Val Pro Asp Thr 220 225 230 235  
Tyr Ser Asn Ala Thr Leu Trp Tyr Lys Ile Phe Thr Thr Ala Arg Asp 240 245 250 255  
Ala Asn Thr Lys Tyr Ala Gln Asp Tyr Asn Pro Phe Trp Cys Tyr Lys 260 265 270 275  
Gly Ala Ile Gly Lys Val Tyr His Ala Leu Asn Pro Lys Leu Thr Val 280 285 290 295



Ile Val Pro Asp Asp Arg Ser Leu Ile Asn Leu His Leu Met His  
260 265 270  
5 Thr Ser Tyr Phe Leu Phe Val Met Val Ile Thr Met Phe Cys Tyr Ala  
275 280 285

Val Ile Lys Gly Arg Pro Ser Lys Leu Arg Gln Ser Asn Pro Glu Phe  
290 295 300

Cys Pro Glu Lys Val Ala Leu Ala Glu Ala Xaa  
305 310 315

(2) INFORMATION FOR SEQ ID NO: 452:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 452:

Met Pro Gly Leu Ser Leu Ala Leu Leu Pro Phe Gly Pro Gly Cys Thr  
1 5 10 15

Glu Ala Leu His Ala Gly Cys Phe Pro Ala Phe Ala Ser Ala Thr Arg  
20 25 30

Val Asn Gly Glu Ala Ala Leu Ser Pro Gly Leu Cys Asp Pro Ile Ser  
35 40 45

Val Pro Tyr Val  
50

35

(2) INFORMATION FOR SEQ ID NO: 453:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 383 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 453:

Met Ala Val Gly Gln Ile Met Thr Phe Gly Ser Pro Val Ile Gly Cys  
1 5 10 15

Gly Phe Ile Ser Gly Trp Asn Leu Val Ser Met Cys Val Gly Tyr Val  
20 25 30

Leu Leu Trp Lys Val Tyr Gln Lys Thr Pro Ala Leu Ala Val Lys Ala  
35 40 45

Gly Leu Lys Glu Glu Thr Glu Leu Lys Gln Leu Asn Leu His Lys  
50 55 60

Asp Thr Glu Pro Lys Pro Leu Glu Gly Thr His Leu Met Gly Val Lys  
65 70 75 80

60

Asp Ser Asn Ile His Glu Leu Glu His Glu Gln Glu Pro Thr Cys Ala  
85 90 95

Ser Gln Met Ala Glu Pro Phe Arg Thr Phe Arg Asp Gly Trp Val Ser  
100 105 110

Tyr Tyr Asn Gln Pro Val Phe Leu Ala Gly Met Gly Leu Ala Phe Leu  
115 120 125

Tyr Met Thr Val Leu Gly Phe Asp Cys Ile Thr Thr Gly Tyr Ala Tyr  
130 135 140

Thr Gln Gly Leu Ser Gly Phe His Pro Gln Tyr Phe Asp Gly Ser Ile  
145 150 155 160

Ser Tyr Asn Trp Asn Asn Gly Asn Cys Ser Phe Tyr Leu Ala Thr Ser  
165 170 175

Lys Met Trp Phe Gly Ser Ala Gly Leu Ile Ser Gly Leu Ala Gln Leu  
180 185 190

Ser Cys Leu Ile Leu Cys Val Ile Ser Val Phe Met Pro Gly Ser Pro  
195 200 205

Leu Asp Leu Ser Val Ser Pro Phe Glu Asp Ile Arg Ser Arg Phe Ile  
210 215 220

Gln Gly Glu Ser Ile Thr Pro Thr Lys Ile Pro Glu Ile Thr Thr Glu  
225 230 235 240

Ile Tyr Met Ser Asn Gly Ser Asn Ser Ala Asn Ile Val Pro Glu Thr  
245 250 255

Ser Pro Glu Ser Val Pro Ile Ile Ser Val Ser Leu Leu Phe Ala Gly  
260 265 270

Val Ile Ala Ala Arg Ile Gly Leu Trp Ser Phe Asp Leu Thr Val Thr  
275 280 285

Gln Leu Leu Gln Glu Asn Val Ile Glu Ser Glu Arg Gly Ile Ile Asn  
290 295 300

Gly Val Gln Asn Ser Met Asn Tyr Leu Leu Asp Leu Leu His Phe Ile  
305 310 315 320

Met Val Ile Leu Ala Pro Asn Pro Glu Ala Phe Gly Leu Leu Val Leu  
325 330 335

Ile Ser Val Ser Phe Val Ala Met Gly His Ile Met Tyr Phe Arg Phe  
340 345 350

Ala Gln Asn Thr Leu Gly Asn Lys Leu Phe Ala Cys Gly Pro Asp Ala  
355 360 365

Lys Glu Val Arg Lys Glu Asn Gln Ala Asn Thr Ser Val Val Xaa  
370 375 380

(2) INFORMATION FOR SEQ ID NO: 454:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 186 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 454:

Met Arg Ser Ile Gly Asn Lys Asn Thr Ile Leu Leu Gly Leu Gly Phe 1 5 10 15  
 Gln Ile Leu Gln Leu Ala Trp Tyr Gly Phe Gly Ser Glu Pro Trp Met 20 25 30  
 Met Trp Ala Ala Gly Ala Val Ala Ala Met Ser Ile Thr Phe Pro 35 40 45  
 Ala Val Ser Ala Leu Val Ser Arg Thr Ala Asp Ala Asp Gln Gln Gly 50 55 60  
 Val Val Gln Gly Met Ile Thr Gly Ile Arg Gly Leu Cys Asn Gly Leu 65 70 75 80  
 Gly Pro Ala Leu Tyr Gly Phe Ile Phe Tyr Ile Phe His Val Glu Leu 85 90 95  
 Lys Glu Leu Pro Ile Thr Gly Thr Asp Leu Gly Thr Asn Thr Ser Pro 100 105 110 115  
 Gln His Phe Glu Gln Asn Ser Ile Ile Pro Gly Pro Pro Phe Leu 120 125  
 Phe Gly Ala Cys Ser Val Leu Leu Ala Leu Leu Val Ala Leu Phe Ile 130 135 140 145  
 Pro Glu His Thr Asn Leu Ser Leu Arg Ser Ser Trp Arg Lys His 150 155 160  
 Cys Gly Ser His Ser His Pro His Asn Thr Gln Ala Pro Gly Glu Ala 165 170 175  
 Lys Glu Pro Leu Leu Gln Asp Thr Asn Val 180 185

(2) INFORMATION FOR SEQ ID NO: 455:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 163 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 455:

Met Leu Gln Thr Ser Asn Tyr Ser Leu Val Leu Ser Leu Gln Phe Leu 1 5 10 15  
 Leu Leu Ser Tyr Asp Leu Phe Val Asn Ser Phe Ser Glu Leu Leu Gln 20 25 30  
 Lys Thr Pro Val Ile Gln Leu Val Leu Phe Ile Ile Gln Asp Ile Ala

Val Leu Phe Asn Ile Ile Ile Phe Leu Met Phe Phe Asn Thr Phe 35 40 45  
 Val Phe Gln Ala Gly Leu Val Asn Leu Leu Phe His Lys Phe Lys Gly 50 55 60  
 Thr Ile Ile Leu Thr Ala Val Tyr Phe Ala Leu Ser Ile Ser Leu His 65 70 75 80  
 Val Trp Val Met Asn Leu Arg Trp Lys Asn Ser Asn Ser Phe Ile Trp 85 90 95  
 Thr Asp Gly Leu Gln Met Leu Phe Val Phe Gln Arg Leu Ala Val 100 105 110 115  
 Leu Tyr Cys Tyr Phe Tyr Lys Arg Thr Ala Val Arg Leu Gly Asp Pro 120 125 130 135 140  
 His Phe Tyr Gln Asp Ser Leu Trp Leu Arg Lys Glu Phe Met Gln Val 145 150 155 160  
 Arg Arg Xaa

(2) INFORMATION FOR SEQ ID NO: 456:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 46 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 456:

Met Arg Ile Gln Val Phe Ile Leu Leu Gly Ala Gly Gly Thr Ser 1 5 10 15  
 Gln Phe Thr Lys Pro Pro Ser Leu Pro Leu Glu Pro Glu Pro Ala Val 20 25 30  
 Glu Ser Ser Pro Thr Glu Thr Ser Glu Gln Ile Arg Glu Lys 35 40 45

(2) INFORMATION FOR SEQ ID NO: 457:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 105 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 457:

Met Ser Tyr Leu Ala Phe Leu Tyr Met Thr Phe Asp Phe Cys Cys Leu 1 5 10 15  
 Tyr Phe Ser Thr Val Tyr Ala Pro Ser Phe Lys Tyr Ile Cys Val His 20 25 30

Thr Asp Thr His Ile Cys Val Cys Val Cys Ile Tyr Leu Ser Ser Val  
35 45  
5 Val Ser Lys Ser Ser Ala Glu Ala Asp Gly Val Leu Gln Pro Arg Arg  
50 55 60  
His Pro Ala Ser Leu Leu Ile Val Phe Ala Thr Ser Ile Ser Glu Ser  
65 70 75 80  
10 Ser Leu Leu Ile Phe Ser Phe Gln Lys Thr Glu Ala Lys Leu Ile Val  
85 90 95  
Phe Ala Val Ser Leu Ala Lys Xaa  
100 105  
15

(2) INFORMATION FOR SEQ ID NO: 458:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 458:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser  
1 5 10 15

30 Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val  
20 25 30

Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro  
35 40 45

35 Ser Ser Glu Leu Leu Glu Gln Leu Leu Ser Val Gln Phe Val Trp Gln  
50 55 60

40 Ala His Thr Val Ala Xaa  
65 70

(2) INFORMATION FOR SEQ ID NO: 459:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 459:

Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Phe Phe  
1 5 10 15

55 Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30

60 Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60

5 Val Gly Phe Leu Glu Leu Ala Gly Leu Leu Val Met Gly Pro  
65 70 75 80

Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95

10 Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110

Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly  
115 120 125

15 Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140

20 Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

(2) INFORMATION FOR SEQ ID NO: 460:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 332 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 460:

Met Lys Leu Gly Arg Ala Val Leu Gly Leu Leu Leu Ala Pro Ser  
1 5 10 15

35 Val Val Gln Ala Val Glu Pro Ile Ser Leu Gly Leu Ala Leu Ala Gly  
20 25 30

40 Val Leu Thr Gly Tyr Ile Tyr Pro Arg Leu Tyr Cys Leu Phe Ala Glu  
35 40 45

Cys Cys Gly Gln Lys Arg Ser Leu Ser Arg Glu Ala Leu Gln Lys Asp  
50 55 60

45 Leu Asp Asp Asn Leu Phe Gly Gln His Leu Ala Lys Ile Ile Leu  
65 70 75 80

Asn Ala Val Phe Gly Phe Ile Asn Asn Pro Lys Pro Lys Lys Pro Leu  
85 90 95

50 Thr Leu Ser Leu His Gly Trp Thr Gly Thr Gly Lys Asn Phe Val Ser  
100 105 110

Lys Ile Ile Ala Glu Asn Ile Tyr Glu Gly Gly Leu Asn Ser Asp Tyr  
115 120 125

55 Val His Leu Phe Val Ala Thr Leu His Phe Pro His Ala Ser Asn Ile  
130 135 140

60 Thr Leu Tyr Lys Asp Gln Leu Gln Leu Trp Ile Arg Gly Asn Val Ser  
145 150 155 160

Ala Cys Ala Arg Ser Ile Phe Ile Phe Asp Glu Met Asp Lys Met His  
165 170 175

5 Ala Gly Leu Ile Asp Ala Ile Lys Pro Phe Leu Asp Tyr Tyr Asp Leu  
180 185 190

Val Asp Gly Val Ser Tyr Gln Lys Ala Met Phe Ile Phe Leu Ser Asn  
195 200 205

10 Ala Gly Ala Glu Arg Ile Thr Asp Val Ala Leu Asp Phe Tyr Arg Ser  
210 215 220

15 Gly Lys Gln Arg Glu Asp Ile Lys Leu Lys Asp Ile Glu His Ala Leu  
225 230 235 240

Ser Val Ser Val Phe Asn Asn Lys Asn Ser Gly Phe Thr His Ser Ser  
245 250 255

20 Leu Ile Asp Arg Asn Leu Ile Asp Tyr Phe Val Pro Phe Leu Pro Leu  
260 265 270

Glu Tyr Lys His Leu Lys Met Cys Ile Arg Val Glu Met Gln Ser Arg  
275 280 285

25 Gly Tyr Glu Ile Asp Glu Asp Ile Val Ser Arg Val Ala Glu Glu Met  
290 295 300

30 Thr Phe Phe Pro Lys Glu Glu Arg Val Phe Ser Asp Lys Gly Cys Lys  
305 310 315 320

Thr Val Phe Thr Lys Leu Asp Tyr Tyr Asp Asp  
325 330

35

(2) INFORMATION FOR SEQ ID NO: 461:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 461:

45 Met Leu Lys Cys Ile  
1 5

50

(2) INFORMATION FOR SEQ ID NO: 462:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 14 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 462:

Met Ile Leu Thr Leu Leu Ser Val Val Ser Thr Met Ala Ser  
1 5 10

60

(2) INFORMATION FOR SEQ ID NO: 463:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 285 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 463:

10 Met Lys Leu His Pro Pro Pro Ser Pro Val Thr Gln Asp His Arg  
1 5 10 15

15 Ser Lys Ser Ser His Ser Asn Trp Met Pro Arg Met Gly Ala Cys Ser  
20 25 30

Met Ser Arg Thr Ser Ser Gly Pro Pro Ser Leu Cys Lys Ser Thr  
35 40 45

20 Ser Gly Arg Ser Cys Thr Arg Pro His Cys Trp Pro Ser Leu Pro Ala  
50 55 60

Trp Val Ser Val Phe Thr Arg Thr Asn Thr Gly Ser Trp Cys Tyr Pro  
65 70 75 80

25 Ala Trp Gly Gly Ala Phe Ser Arg Pro Trp Met Ser Ala Gln Ser Met  
85 90 95

30 Cys Cys Ala Glu Arg Ser Val Leu Gln Val Ala Cys Arg Leu Leu Asp  
100 105 110

Ala Leu Glu Phe Leu His Glu Asn Glu Tyr Val His Gly Asn Val Thr  
115 120 125

35 Ala Glu Asn Ile Phe Val Asp Pro Glu Asp Gln Ser Gln Val Thr Leu  
130 135 140

Ala Gly Tyr Gly Phe Ala Phe Arg Tyr Cys Pro Ser Gly Lys His Val  
145 150 155 160

40 Ala Tyr Val Glu Gly Ser Arg Ser Pro His Glu Gly Asp Leu Glu Phe  
165 170 175

45 Ile Ser Met Asp Leu His Lys Gly Cys Gly Pro Ser Arg Arg Xaa Asp  
180 185 190

Leu Gln Ser Leu Gly Tyr Cys Met Leu Lys Trp Leu Tyr Gly Phe Leu  
195 200 205

50 Pro Trp Thr Asn Cys Leu Pro Xaa Xaa Glu Asp Ile Met Lys Gln Lys  
210 215 220

Gln Lys Phe Val Asp Lys Pro Gly Pro Phe Val Gly Pro Cys Gly His  
225 230 235 240

55 Trp Ile Arg Pro Ser Glu Thr Leu Gln Lys Tyr Leu Lys Val Val Met  
245 250 255

60 Ala Leu Thr Tyr Glu Glu Lys Pro Pro Tyr Ala Met Leu Arg Asn  
260 265 270

632

633

Leu Glu Ala Leu Leu Gln Asp Leu Arg Val Ser Pro Tyr  
275 280 285

5

(2) INFORMATION FOR SEQ ID NO: 464:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 464:

Met Thr Ser Pro Pro His Gln Gly Trp Glu Gln Arg Gly Cys Gly  
1 5 10 15

Glu Ser Gln Val Pro Leu Ala Leu Ser Arg Val Phe Ser Thr Ser His  
20 25 30

Tyr Cys Leu Leu Leu Val Ala Asn Gln Ser Ile Phe Pro Cys Leu  
35 40 45

Trp Ala Val Glu Arg Leu Leu Gly Val Arg Cys Thr Cys Pro Leu Ser  
50 55 60

Trp Gly Lys Arg Ile Ile Ser Glu His Cys Ser Ala Gln Ser Ser Xaa  
65 70 75 80

30

(2) INFORMATION FOR SEQ ID NO: 465:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 465:

Met His Thr Trp Tyr Asn Asp Arg Arg Gln Asn Cys His Cys Leu Leu  
1 5 10 15

Phe Phe Leu Ile Tyr Leu Arg Lys Ile Tyr Gln Val Val Pro His Val  
20 25 30

Pro Leu Leu Val Lys Cys Arg Gly Arg Leu Lys Gly Val Asn Ile  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 466:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 96 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 466:

633

Met Glu Leu Val Leu Val Phe Leu Cys Ser Leu Leu Ala Pro Met Val  
1 5 10 15

Leu Ala Ser Ala Ala Glu Lys Glu Met Asp Pro Phe His Tyr  
20 25 30

Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu Val Phe Ala Val Leu  
35 40 45

Phe Ser Val Gly Ile Leu Leu Ile Leu Ser Arg Arg Cys Lys Cys Ser  
50 55 60

Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp Glu Glu Ala Gln Val Glu  
65 70 75 80

Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro Gln Lys Ala Glu Asn Xaa  
85 90 95

20

(2) INFORMATION FOR SEQ ID NO: 467:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 399 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 467:

Met Ala Ser Gly Ala Asp Ser Lys Gly Asp Asp Leu Ser Thr Ala Ile  
1 5 10 15

Leu Lys Gln Lys Asn Arg Pro Asn Arg Leu Ile Val Asp Glu Ala Ile  
20 25 30

Asn Glu Asp Asn Ser Val Val Ser Leu Ser Gln Pro Lys Met Asp Glu  
35 40 45

Leu Gln Leu Phe Arg Gly Asp Thr Val Leu Leu Lys Gly Lys Lys Arg  
50 55 60

Arg Glu Ala Val Cys Ile Val Leu Ser Asp Asp Thr Cys Ser Asp Glu  
65 70 75 80

Lys Ile Arg Met Asn Arg Val Val Arg Asn Asn Leu Arg Val Arg Leu  
85 90 95

Gly Asp Val Ile Ser Ile Gln Pro Cys Pro Asp Val Lys Tyr Gly Lys  
100 105 110

Arg Ile His Val Leu Pro Ile Asp Asp Thr Val Glu Gly Ile Thr Gly  
115 120 125

Asn Leu Phe Glu Val Tyr Leu Lys Pro Tyr Phe Leu Glu Ala Tyr Arg  
130 135 140

Pro Ile Arg Lys Gly Asp Ile Phe Leu Val Arg Gly Gly Met Arg Ala  
145 150 155

634

145 150 155 160  
Val Glu Phe Lys Val Val Glu Thr Asp Pro Ser Pro Tyr Cys Ile Val 175  
165  
5 Ala Pro Asp Thr Val Ile His Cys Glu Glu Pro Ile Lys Arg Glu 190  
185  
Asp Glu Glu Glu Ser Leu Asn Glu Val Gly Tyr Asp Asp Ile Gly Gly 205  
195 200  
10 Cys Arg Lys Glu Leu Ala Gln Ile Lys Glu Met Val Glu Leu Pro Leu 220  
210 215  
15 Arg His Pro Ala Leu Phe Lys Ala Ile Gly Val Lys Pro Pro Arg Gly 240  
225 230 235  
Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly Lys Thr Leu Ile Ala Arg 255  
245 250  
20 Ala Val Ala Asn Glu Thr Gly Ala Phe Phe Leu Ile Asn Gly Pro 270  
260 265  
Glu Ile Met Ser Lys Leu Ala Gly Glu Ser Glu Ser Asn Leu Arg Lys 285  
275 280  
25 Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro Ala Ile Ile Phe Ile Asp 300  
290 295  
30 Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu Lys Thr His Gly Glu Val 320  
305 310 315  
Glu Arg Arg Ile Val Ser Gln Leu Leu Thr Leu Met Asp Gly Leu Lys 335  
325 330  
35 Gln Arg Ala His Val Ile Val Met Ala Ala Thr Asn Arg Pro Asn Ser 350  
340 345  
Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg Phe Asp Arg Glu Val Asp 365  
355 360  
40 Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu Glu Ile Leu Gln Ile His 380  
370 375  
45 Thr Lys Asn Met Lys Leu Ala Asp Asp Val Asp Leu Glu Gln Xaa 395  
385 390 395

(2) INFORMATION FOR SEQ ID NO: 468:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 468:

Leu 1

60

635

(2) INFORMATION FOR SEQ ID NO: 469:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 273 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 469:

Met Ala Ala Pro Lys Gly Ser Leu Trp Val Arg Thr Gln Leu Gly Leu 15  
1 5 10  
10 Pro Pro Leu Leu Leu Thr Met Ala Leu Ala Gly Gly Ser Gly Thr 30  
20 25  
15 Ala Ser Ala Glu Ala Phe Asp Ser Val Leu Gly Asp Thr Ala Ser Cys 45  
35 40  
20 His Arg Ala Cys Gln Leu Thr Tyr Pro Leu His Thr Tyr Pro Lys Glu 60  
50 55  
Glu Glu Leu Tyr Ala Cys Gln Arg Gly Cys Arg Leu Phe Ser Ile Cys 80  
65 70 75  
25 Gln Phe Val Asp Asp Gly Ile Asp Leu Asn Arg Thr Lys Leu Glu Cys 95  
85 90  
30 Glu Ser Ala Cys Thr Gln Ala Tyr Ser Gln Ser Asp Glu Gln Tyr Ala 110  
100 105  
Cys His Leu Gly Cys Gln Asn Gln Leu Pro Phe Ala Glu Leu Arg Gln 125  
115 120  
35 Glu Gln Leu Met Ser Leu Met Pro Lys Met His Leu Leu Phe Pro Leu 140  
130 135  
Thr Leu Val Arg Ser Phe Thr Ser Asp Met Met Asp Ser Ala Gln Ser 160  
145 150 155  
40 Phe Ile Thr Ser Ser Thr Thr Phe Tyr Leu Gln Ala Asp Asp Gly Lys 175  
165 170  
45 Ile Val Ile Phe Xaa Ser Lys Pro Arg Asn Pro Arg Tyr Ala Pro His 190  
180 185  
Leu Glu Pro Gly Ala Leu Pro Asn Leu Xaa Xaa Ser Leu Ser Lys 205  
195 200  
50 Met Ser Xaa Xaa Ser Xaa Met Arg Asn Ser Gln Ala His Arg Asn Phe 220  
210 215  
Leu Glu Asp Gly Glu Ser Asp Gly Phe Leu Arg Cys Leu Ser Leu Asn 240  
225 230 235  
55 Ser Gly Trp Ile Leu Thr Thr Thr Leu Val Leu Ser Val Met Val Leu 255  
245 250  
60 Leu Trp Ile Cys Cys Ala Thr Cys Cys Tyr Thr Leu Leu Asp Ala Val 270  
260 265

636

637

Xaa

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 471:

5 Met Arg Lys Thr Arg Leu Trp Gly Leu Leu Trp Met Leu Phe Val Ser 15  
1 5 10  
Glu Leu Arg Ala Ala Thr Lys Leu Thr Glu Glu Lys Tyr Glu Leu Lys 30  
20 25  
10 Glu Gly Gln Thr Leu Asp Val Lys Cys Asp Tyr Thr Leu Glu Lys Phe 45  
35 40  
Ala Ser Ser Gln Lys Ala Trp Gln Ile Ile Arg Asp Gly Glu Met Pro 60  
50 55  
15 Lys Thr Leu Ala Cys Thr Glu Arg Pro Ser Lys Asn Ser His Pro Val 80  
65 70 75  
Gln Val Gly Arg Ile Ile Leu Glu Asp Tyr His Asp His Gly Leu Leu 95  
85 90  
Arg Val Arg Met Val Asn Leu Gln Val Glu Asp Ser Gly Leu Tyr Gln 110  
100 105  
25 Cys Val Ile Tyr Gln Pro Pro Lys Glu Pro His Met Leu Phe Asp Arg 125  
115 120  
Ile Arg Leu Val Val Thr Lys Gly Phe Ser Gly Thr Pro Gly Ser Asn 140  
130 135  
30 Glu Asn Ser Thr Gln Asn Val Tyr Lys Ile Pro Pro Thr Thr Thr Lys 160  
145 150 155  
Ala Leu Cys Pro Leu Tyr Thr Ser Pro Arg Thr Val Thr Gln Ala Pro 175  
165 170  
35 Pro Lys Ser Thr Ala Asp Val Ser Thr Pro Asp Ser Glu Ile Asn Leu 190  
180 185  
40 Thr Asn Val Thr Asp Ile Ile Arg Val Pro Val Phe Asn Ile Val Ile 205  
195 200  
Leu Leu Ala Gly Phe Leu Ser Lys Ser Leu Val Phe Ser Val Leu 220  
210 215  
45 Phe Ala Val Thr Leu Arg Ser Phe Val Pro 230  
225 235  
50

(2) INFORMATION FOR SEQ ID NO: 472:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 105 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 472:

60 Met Leu His Ile Leu Pro Leu Lys Ser Tyr Asp Phe Pro His Phe Ser 15  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 470:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 192 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(X1) SEQUENCE DESCRIPTION: SEQ ID NO: 470:

15 Met Met Val Leu Ser Leu Gly Ile Ile Leu Ala Ser Ala Ser Phe Ser 15  
1 5 10  
Pro Asn Phe Thr Gln Val Thr Ser Thr Leu Leu Asn Ser Ala Tyr Pro 30  
20 25  
Phe Ile Gly Pro Phe Phe Ile Ile Ser Gly Ser Leu Ser Ile Ala 45  
35 40  
Thr Glu Lys Arg Leu Thr Lys Leu Val His Ser Ser Leu Val Gly 60  
50 55  
25 Ser Ile Leu Ser Ala Leu Ser Ala Leu Val Gly Phe Ile Ile Leu Ser 80  
65 70 75  
30 Val Lys Gln Ala Thr Leu Asn Pro Ala Ser Leu Gln Cys Glu Leu Asp 95  
85 90  
Lys Asn Asn Ile Pro Thr Arg Ser Tyr Val Ser Tyr Phe Tyr His Asp 110  
100 105  
35 Ser Leu Tyr Thr Thr Asp Cys Tyr Thr Ala Lys Ala Ser Leu Ala Gly 125  
115 120  
40 Xaa Leu Ser Leu Met Leu Ile Cys Thr Leu Leu Glu Phe Cys Leu Ala 140  
130 135  
Val Leu Thr Ala Val Leu Arg Trp Lys Gln Ala Tyr Ser Asp Phe Pro 160  
145 150 155  
45 Gly Ser Val Leu Phe Leu Pro His Ser Tyr Tyr Ile Gly Asn Ser Gly Met 175  
165 170  
Ser Ser Lys Met Thr His Asp Cys Gly Tyr Glu Glu Leu Thr Ser 190  
180 185  
50

(2) INFORMATION FOR SEQ ID NO: 471:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 234 amino acids

(B) TYPE: amino acid

Leu Met Gly Arg Tyr Arg Cys Ala Ser Leu Leu Phe Cys Phe Leu Leu  
 20 25 30  
 5 Leu Phe Phe Phe Cys Ser Val Leu Trp Thr Phe Ser Asp Met His  
 35 40 45  
 Arg Ser Gly Glu Asp Gly Pro Trp Thr Pro Cys Val His His Leu Ala  
 50 55 60  
 10 Ala Ser Leu Ile Ser Tyr Gly Gln Pro Gly Phe Ile Cys Ile Ser Leu  
 65 70 75 80  
 Phe Ser Pro Val Leu Phe Ile Glu Asn Pro Arg His Tyr Ala Asn Ala  
 85 90 95  
 Thr Val Thr Thr Leu Gly Asp Trp Xaa  
 100 105

## (2) INFORMATION FOR SEQ ID NO: 473:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 32 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 473:

Met Val Phe Leu Lys Tyr Arg Phe Phe Leu Val Phe Leu Ala  
 1 5 10 15  
 30 Asn Cys Ile Tyr Ser Leu His Tyr Lys Pro Ser Leu Met Tyr Pro Lys  
 20 25 30

## (2) INFORMATION FOR SEQ ID NO: 474:

- (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 571 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 474:

Met Ala Leu Ser Arg Gly Leu Pro Arg Glu Leu Ala Glu Ala Val Ala  
 1 5 10 15  
 Gly Gly Arg Val Leu Val Val Gly Ala Gly Gly Ile Gly Cys Glu Leu  
 20 25 30  
 55 Leu Lys Asn Leu Val Leu Thr Gly Phe Ser His Ile Asp Leu Ile Asp  
 35 40 45  
 Leu Asp Thr Ile Asp Val Ser Asn Leu Asn Arg Gln Phe Leu Phe Gln  
 50 55 60

Lys Lys His Val Gly Arg Ser Lys Ala Gln Val Ala Lys Glu Ser Val  
 65 70 75 80  
 5 Leu Gln Phe Tyr Pro Lys Ala Asn Ile Val Ala Tyr His Asp Ser Ile  
 85 90 95  
 Met Asn Pro Asp Tyr Asn Val Glu Phe Arg Gln Phe Ile Leu Val  
 100 105 110  
 10 Met Asn Ala Leu Asp Asn Arg Ala Ala Arg Asn His Val Asn Arg Met  
 115 120 125  
 Cys Leu Ala Ala Asp Val Pro Leu Ile Glu Ser Gly Thr Ala Gly Tyr  
 130 135 140  
 15 Leu Gly Gln Val Thr Thr Ile Lys Lys Gly Val Thr Glu Cys Tyr Glu  
 145 150 155 160  
 Cys His Pro Lys Pro Thr Gln Arg Thr Phe Pro Gly Cys Thr Ile Arg  
 165 170 175  
 20 Asn Thr Pro Ser Glu Pro Ile His Cys Ile Val Trp Ala Lys Tyr Leu  
 180 185 190  
 25 Phe Asn Gln Leu Phe Gly Glu Asp Ala Asp Gln Glu Val Ser Pro  
 195 200 205  
 Asp Arg Ala Asp Pro Glu Ala Ala Trp Glu Pro Thr Glu Ala Glu Ala  
 210 215 220  
 30 Arg Ala Arg Ala Ser Asn Glu Asp Gly Asp Ile Lys Arg Ile Ser Thr  
 225 230 235  
 35 Lys Glu Trp Ala Lys Ser Thr Gly Tyr Asp Pro Val Lys Leu Phe Thr  
 240 245 250 255  
 Lys Leu Phe Lys Asp Asp Ile Arg Tyr Leu Leu Thr Met Asp Lys Leu  
 260 265 270  
 40 Trp Arg Lys Arg Lys Pro Pro Val Pro Leu Asp Trp Ala Glu Val Gln  
 275 280 285  
 Ser Gln Gly Glu Thr Asn Ala Ser Asp Gln Gln Asn Glu Pro Gln  
 290 295 300  
 45 Leu Gly Leu Lys Asp Gln Gln Val Leu Asp Val Lys Ser Tyr Ala Arg  
 305 310 315 320  
 Leu Phe Ser Lys Ser Ile Glu Thr Leu Arg Val His Leu Ala Glu Lys  
 325 330 335  
 50 Gly Asp Gly Ala Glu Leu Ile Trp Asp Lys Asp Asp Pro Ser Ala Met  
 340 345 350  
 55 Asp Phe Val Thr Ser Ala Ala Asn Leu Arg Met His Ile Phe Ser Met  
 355 360 365  
 Asn Met Lys Ser Arg Phe Asp Ile Lys Ser Met Ala Gly Asn Ile Ile  
 370 375 380  
 60



Pro Ala Ile Ala Thr Thr Asn Ala Val Ile Ala Gly Leu Ile Val Leu  
385 390 395 400

Glu Gly Leu Lys Ile Leu Ser Gly Lys Ile Asp Gln Cys Arg Thr Ile  
405 410 415

Phe Leu Asn Lys Gln Pro Asn Pro Arg Lys Lys Leu Leu Val Pro Cys  
420 425 430

Ala Leu Asp Pro Pro Asn Pro Asn Cys Tyr Val Cys Ala Ser Lys Pro  
435 440 445

Glu Val Thr Val Arg Leu Asn Val His Lys Val Thr Val Leu Thr Leu  
450 455 460

Gln Asp Lys Ile Val Lys Glu Lys Phe Ala Met Val Ala Pro Asp Val  
465 470 475 480

Gln Ile Glu Asp Gly Lys Gly Thr Ile Leu Ile Ser Ser Glu Glu Gly  
485 490 495

Glu Thr Glu Ala Asn Asn His Lys Lys Leu Ser Glu Phe Gly Ile Arg  
500 505 510

Asn Gly Ser Arg Leu Gln Ala Asp Asp Phe Leu Gln Asp Tyr Thr Leu  
515 520 525

Leu Ile Asn Ile Leu His Ser Glu Asp Leu Gly Lys Asp Val Glu Phe  
530 535 540

Glu Val Val Gly Asp Ala Pro Glu Lys Val Gly Xaa Lys Gln Ala Glu  
545 550 555 560

Asp Ala Ala Lys Ser Ile Thr Asn Gly Gln Xaa  
565 570

(2) INFORMATION FOR SEQ ID NO: 475:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 312 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 475:

Met Gln Val Val Thr Cys Leu Thr Arg Asp Ser Tyr Leu Thr His Cys  
1 5 10 15

Phe Leu Gln His Leu Met Val Val Leu Ser Ser Leu Glu Arg Thr Pro  
20 25 30

Ser Pro Glu Pro Val Asp Lys Asp Phe Tyr Ser Glu Phe Gly Asn Lys  
35 40 45

Thr Thr Gly Lys Met Glu Asn Tyr Glu Leu Ile His Ser Ser Arg Val  
50 55 60

Lys Phe Thr Tyr Pro Ser Glu Glu Ile Gly Asp Leu Thr Phe Thr  
65 70 75 80

Val Ala Gln Lys Met Ala Glu Pro Glu Lys Ala Pro Ala Leu Ser Ile  
85 90 95

Leu Leu Tyr Val Gln Ala Phe Gln Val Gly Met Pro Pro Gly Cys  
100 105 110

Cys Arg Gly Pro Leu Arg Pro Lys Thr Leu Leu Thr Ser Ser Glu  
115 120 125

Ile Phe Leu Leu Asp Glu Asp Cys Val His Tyr Pro Leu Pro Glu Phe  
130 135 140

Ala Lys Glu Pro Pro Gln Arg Asp Tyr Arg Leu Asp Asp Gly Arg  
145 150 155 160

Arg Val Arg Asp Leu Asp Arg Val Leu Met Gly Tyr Gln Thr Tyr Pro  
165 170 175

Gln Pro Ser Pro Ser Ser Met Thr Cys Lys Val Met Thr Ser Trp  
180 185 190

Ala Val Ser Pro Trp Thr Thr Leu Gly Arg Cys Gln Val Ala Arg Leu  
195 200 205

Glu Pro Ala Arg Ala Val Lys Ser Ser Gly Arg Cys Leu Ser Pro Val  
210 215 220

Leu Arg Ala Glu Arg Ser Ser Arg Cys Trp Leu Ala Ser Gly Arg  
225 230 235 240

Pro Cys Val Ala Val Ser Cys Leu Ser Ser Pro Ala Ser Pro Gly  
245 250 255

His Ser Gln Pro Val Val Ser Ser Leu Thr Pro Thr Gly Ala Gly Gln  
260 265 270

Gln Ala Phe Val Phe Ser Lys Asn Val Leu Ser Ser Leu Trp Tyr Leu  
275 280 285

Asn Leu Thr Val Leu Ala Glu Asn Val Asn Met Cys Val Cys Cys Val  
290 295 300 305

Asn Ser Phe Ser Cys Trp Glu Xaa  
310

(2) INFORMATION FOR SEQ ID NO: 476:

(i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 329 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 476:

Met Ala Gln His His Leu Trp Ile Leu Leu Leu Cys Leu Gln Thr Trp  
1 5 10 15

Pro Glu Ala Ala Gly Lys Asp Ser Glu Ile Phe Thr Val Asn Gly Ile

20 25 30  
 Leu Gly Glu Ser Val Thr Phe Pro Val Asn Ile Gln Glu Pro Arg Gln 45  
 35  
 5 Val Lys Ile Ile Ala Trp Thr Ser Lys Thr Ser Val Ala Tyr Val Thr 60  
 50  
 Pro Gly Asp Ser Glu Thr Ala Pro Val Val Thr Val Thr His Arg Asn 80  
 65  
 10 Tyr Tyr Glu Arg Ile His Ala Leu Gly Pro Asn Tyr Asn Leu Val Ile 95  
 85  
 Ser Asp Leu Arg Met Glu Asp Ala Gly Asp Tyr Lys Ala Asp Ile Asn 110  
 100  
 Thr Gln Ala Asp Pro Tyr Thr Thr Thr Lys Arg Tyr Asn Leu Gln Ile 125  
 115  
 20 Tyr Arg Arg Leu Gly Lys Pro Lys Ile Thr Gln Ser Leu Met Ala Ser 140  
 130  
 Val Asn Ser Thr Cys Asn Val Thr Leu Thr Cys Ser Val Glu Lys Glu 160  
 145  
 25 Glu Lys Asn Val Thr Tyr Asn Trp Ser Pro Leu Gly Glu Gly Asn 175  
 165  
 30 Val Leu Gln Ile Phe Gln Thr Pro Glu Asp Gln Glu Leu Thr Tyr Thr 190  
 180  
 Cys Thr Ala Gln Asn Pro Val Ser Asn Asn Ser Asp Ser Ile Ser Ala 205  
 195  
 35 Arg Gln Leu Cys Ala Asp Ile Ala Met Gly Phe Arg Thr His His Thr 220  
 210  
 Gly Leu Leu Ser Val Leu Ala Met Phe Phe Leu Leu Val Leu Ile Leu 240  
 225  
 Ser Ser Val Phe Leu Phe Arg Leu Phe Lys Arg Arg Gln Asp Ala Ala 255  
 245  
 40 Ser Lys Lys Thr Ile Tyr Thr Tyr Ile Met Ala Ser Arg Asn Thr Gln 270  
 260  
 Pro Ala Glu Ser Arg Ile Tyr Asp Glu Ile Leu Gln Ser Lys Val Leu 285  
 275  
 50 Pro Ser Lys Glu Glu Pro Val Asn Thr Val Tyr Ser Gln Val Gln Phe 300  
 290  
 Ala Asp Lys Met Gly Lys Ala Ser Thr Gln Asp Ser Lys Pro Pro Gly 320  
 305  
 55 Thr Ser Ser Tyr Glu Ile Val Ile Xaa 325  
 315

(2) INFORMATION FOR SEQ ID NO: 477:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 178 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 477:  
 5 Met Lys Leu Gln Cys Val Ser Leu Trp Leu Leu Gly Thr Ile Leu Ile 15  
 10 Leu Cys Ser Val Asp Asn His Gly Leu Arg Arg Cys Leu Ile Ser Thr 30  
 20 Asp Met His His Ile Glu Glu Ser Phe Gln Glu Ile Lys Arg Ala Ile 45  
 35 Gln Ala Lys Asp Thr Phe Pro Asn Val Thr Ile Leu Ser Thr Leu Glu 60  
 50 Thr Leu Gln Ile Ile Lys Pro Leu Asp Val Cys Cys Val Thr Lys Asn 80  
 65 Leu Leu Ala Phe Tyr Val Asp Arg Val Phe Lys Asp His Gln Glu Pro 95  
 85 Asn Pro Lys Ile Leu Arg Lys Ile Ser Ser Ile Ala Asn Ser Phe Leu 110  
 100 Tyr Met Gln Lys Thr Leu Arg Gln Cys Gln Glu Gln Arg Cys His 125  
 115 Cys Arg Gln Glu Ala Thr Asn Ala Thr Arg Val Ile His Asp Asn Tyr 140  
 130 Asp Gln Leu Glu Val His Ala Ala Ile Lys Ser Leu Gly Glu Leu 160  
 145 Asp Val Phe Leu Ala Trp Ile Asn Lys Asn His Glu Val Met Ser Ser 175  
 165  
 40 Ala Xaa  
 45  
 (2) INFORMATION FOR SEQ ID NO: 478:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 52 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 478:  
 50 Asp Thr Ala Ile Arg Val Ala Leu Ala Val Ala Val Leu Lys Thr Val 15  
 10 Ile Leu Gly Leu Leu Cys Leu Leu Cys Gly Gly Glu Gly Lys 30  
 25

Val Ala Gly Arg Gln Ala Val Thr Ser Asp Gln Gln Ser Val Gly Arg  
35 40 45

5 Arg Asp Val Tyr  
50

10 (2) INFORMATION FOR SEQ ID NO: 479:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 62 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 479:

Met Gln Lys Lys Asn Ser Leu Phe Phe Phe Ala Phe Tyr Tyr Glu  
1 5 10 15

20 Asn Lys Thr Asn Ala Pro Gly Glu Gly Ser Met Ile Thr Arg Asn Ile  
20 25 30

25 Lys Glu Tyr Phe Leu Pro Phe Leu Phe Cys Cys Val Glu Ala Ser Ile  
35 40 45

Ala Ile Asn Lys Leu Asn Tyr Leu His Trp Thr His Phe Gln  
50 60

30

(2) INFORMATION FOR SEQ ID NO: 480:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 480:

40 Met Pro Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15

Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

45

(2) INFORMATION FOR SEQ ID NO: 481:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 339 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 481:

Met Ser Gly Pro Asp Val Glu Thr Pro Ser Ala Ile Gln Ile Cys Arg  
1 5 10 15

Ile Met Arg Pro Asp Asp Ala Asn Val Ala Gly Asn Val His Gly Gly  
20 25 30

60

Thr Ile Leu Lys Met Ile Glu Glu Ala Gly Ala Ile Ile Ser Thr Arg  
35 40 45

5 His Cys Asn Ser Gln Asn Gly Glu Arg Cys Val Ala Ala Leu Ala Arg  
50 55 60

10 Val Glu Arg Thr Asp Phe Leu Ser Pro Met Cys Ile Gly Glu Val Ala  
65 70 75 80

His Val Ser Ala Glu Ile Thr Tyr Thr Ser Lys His Ser Val Glu Val  
85 90 95

15 Gln Val Asn Val Met Ser Glu Asn Ile Leu Thr Gly Ala Lys Lys Leu  
100 105 110

Thr Asn Lys Ala Thr Leu Trp Tyr Val Pro Leu Ser Leu Lys Asn Val  
115 120 125

20 Asp Lys Val Leu Glu Val Pro Val Tyr Ser Arg Xaa Glu Gln  
130 135 140

Glu Glu Glu Gly Arg Lys Arg Tyr Glu Ala Gln Lys Leu Glu Arg Met  
145 150 155 160

25 Glu Thr Lys Trp Arg Asn Gly Asp Ile Val Gln Pro Val Leu Asn Pro  
165 170 175

30 Glu Pro Asn Thr Val Ser Tyr Ser Gln Ser Ser Leu Ile His Leu Val  
180 185 190

Gly Pro Ser Asp Cys Thr Leu His Gly Phe Val His Gly Gly Val Thr  
195 200 205

35 Met Lys Leu Met Asp Glu Val Ala Gly Ile Val Ala Ala Arg His Cys  
210 215 220

Lys Thr Asn Ile Val Thr Ala Ser Val Asp Ala Ile Asn Phe His Asp  
225 230 235 240

40 Lys Ile Arg Lys Gly Cys Val Ile Thr Ile Ser Gly Arg Met Thr Phe  
245 250 255

Thr Ser Asn Lys Ser Met Glu Ile Glu Val Leu Val Asp Ala Asp Pro  
260 265 270

45 Val Val Asp Ser Ser Gln Lys Arg Tyr Arg Ala Ala Ser Ala Phe Phe  
275 280 285

50 Thr Tyr Val Ser Leu Ser Gln Glu Gly Arg Ser Leu Pro Val Pro Gln  
290 295 300

Leu Val Pro Glu Thr Glu Asp Glu Lys Arg Phe Glu Glu Gly Lys  
305 310 315 320

55 Gly Arg Tyr Leu Gln Met Lys Ala Lys Xaa Gln Gly His Ala Xaa Xaa  
325 330 335

Gln Pro Xaa

60

646

647

## (2) INFORMATION FOR SEQ ID NO: 482:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 482:

Met Leu Asn Ser Asn Ile Asn Asp Leu Leu Met Val Thr Tyr Leu Ala  
1 5 10 15

Asn Leu Thr Gln Ser Gln Ile Ala Leu Asn Glu Lys Leu Val Asn Leu  
20 25 30

20

## (2) INFORMATION FOR SEQ ID NO: 483:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 48 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 483:

Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr Arg Ile Gly  
20 25 30

Cys Phe Lys Thr Ile Thr Cys Trp Pro Thr Ser Leu Thr Gln Arg Xaa  
35 40 45

40

## (2) INFORMATION FOR SEQ ID NO: 484:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 38 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 484:

Met Tyr Met Tyr Ser Leu Asn Val Phe Leu Ser Phe Ile Phe Leu Ala  
1 5 10 15

Leu Val Phe Lys Cys Val His Val Cys Gln Gly Ala Asn Ala Phe Leu  
20 25 30

Phe Leu Lys Leu Val Phe  
35

60

## (2) INFORMATION FOR SEQ ID NO: 485:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 485:

Met Gly Leu Arg Leu Ile Cys Leu Glu Leu Thr Met Val Lys Ala Leu  
1 5 10 15

Val Cys Glu Met Phe Leu Phe Leu Met Thr Gln Lys Leu Ile Trp  
20 25 30

Gln Glu Cys Thr Glu Lys Phe Ala Lys Leu Leu Val Gln Leu Ile Ser  
35 40 45

Leu Val Phe Ala Trp Glu Phe Ser Glu Asp Thr Pro  
50 55 60

## (2) INFORMATION FOR SEQ ID NO: 486:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 346 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 486:

Met Leu Ala Ala Arg Leu Val Cys Leu Arg Thr Leu Pro Ser Arg Val  
1 5 10 15

Phe His Pro Ala Phe Thr Lys Ala Ser Pro Val Val Lys Asn Ser Ile  
20 25 30

Thr Lys Asn Gln Trp Leu Leu Thr Pro Ser Arg Glu Tyr Ala Thr Lys  
35 40 45

Thr Arg Ile Gly Ile Arg Arg Gly Thr Gly Gln Glu Leu Lys Glu  
50 55 60

Ala Ala Leu Glu Pro Ser Met Glu Lys Ile Phe Lys Ile Asp Gln Met  
65 70 75 80

Gly Arg Trp Phe Val Ala Gly Gly Ala Ala Val Gly Leu Gly Ala Leu  
85 90 95

Cys Tyr Tyr Gly Leu Gly Leu Ser Asn Glu Ile Gly Ala Ile Glu Lys  
100 105 110 115

Ala Val Ile Trp Pro Gln Tyr Val Lys Asp Arg Ile His Ser Thr Tyr  
120 125

Met Tyr Leu Ala Gly Ser Ile Gly Leu Thr Ala Leu Ser Ala Ile Ala  
130 135 140

60

Ile Ser Arg Thr Pro Val Leu Met Asn Phe Met Met Arg Gly Ser Trp  
 145 150 155 160  
 Val Thr Ile Gly Val Thr Phe Ala Ala Met Val Gly Ala Gly Met Leu  
 165 170 175  
 Val Arg Ser Ile Pro Tyr Asp Gln Ser Pro Gly Pro Lys His Leu Ala  
 180 185 190  
 Trp Leu Leu His Ser Gly Val Met Gly Ala Val Val Ala Pro Leu Thr  
 195 200 205  
 Ile Leu Gly Gly Pro Leu Leu Ile Arg Ala Ala Trp Tyr Thr Ala Gly  
 210 215 220  
 Ile Val Gly Gly Leu Ser Thr Val Ala Met Cys Ala Pro Ser Glu Lys  
 225 230 235 240  
 Phe Leu Asn Met Gly Ala Pro Leu Gly Val Gly Leu Gly Leu Val Phe  
 245 250 255  
 Val Ser Ser Leu Gly Ser Met Phe Leu Pro Pro Thr Thr Val Ala Gly  
 260 265 270  
 Ala Thr Tyr Ser Val Ala Met Tyr Gly Gly Leu Val Leu Phe Ser  
 275 280 285  
 Met Phe Leu Leu Tyr Asp Thr Gln Lys Val Ile Lys Arg Ala Glu Val  
 290 295 300  
 Ser Pro Met Tyr Gly Val Gln Lys Tyr Asp Pro Ile Asn Ser Met Leu  
 305 310 315 320  
 Ser Ile Tyr Met Asp Thr Leu Asn Ile Phe Met Arg Val Ala Thr Met  
 325 330 335  
 Leu Ala Thr Gly Gly Asn Arg Lys Lys Xaa.  
 340 345

(2) INFORMATION FOR SEQ ID NO: 487:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

Met Glu Glu Val Leu Leu Leu Gly Leu Lys Asp Arg Glu Gly Tyr Thr  
 1 5 10 15  
 Ser Phe Trp Asn Asp Cys Ile Ser Ser Gly Leu Arg Gly Cys Met Leu  
 20 25 30  
 Ile Glu Leu Ala Leu Arg Gly Arg Leu Gln Leu Glu Ala Cys Gly Met  
 35 40 45  
 Arg Arg Lys Ser Leu Leu Thr Arg Lys Val Ile Cys Lys Ser Asp Ala  
 50 55 60

(2) INFORMATION FOR SEQ ID NO: 487:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 237 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 487:

Met Glu Glu Val Leu Leu Leu Gly Leu Lys Asp Arg Glu Gly Tyr Thr  
 1 5 10 15  
 Ser Phe Trp Asn Asp Cys Ile Ser Ser Gly Leu Arg Gly Cys Met Leu  
 20 25 30  
 Ile Glu Leu Ala Leu Arg Gly Arg Leu Gln Leu Glu Ala Cys Gly Met  
 35 40 45  
 Arg Arg Lys Ser Leu Leu Thr Arg Lys Val Ile Cys Lys Ser Asp Ala  
 50 55 60

Pro Thr Gly Asp Val Leu Leu Asp Glu Ala Leu Lys His Val Lys Glu  
 65 70 75 80  
 Thr Gln Pro Pro Glu Thr Val Gln Asn Trp Ile Glu Leu Leu Ser Gly  
 85 90 95  
 Glu Thr Trp Asn Pro Leu Lys Leu His Tyr Gln Leu Arg Asn Val Arg  
 100 105 110  
 Glu Arg Leu Ala Lys Asn Leu Val Glu Lys Gly Val Leu Thr Thr Glu  
 115 120 125  
 Lys Gln Asn Phe Leu Phe Asp Met Thr Thr His Pro Leu Thr Asn  
 130 135 140  
 Asn Asn Ile Lys Gln Arg Leu Ile Lys Lys Val Gln Glu Ala Val Leu  
 145 150 155 160  
 Asp Lys Trp Val Asn Asp Pro His Arg Met Asp Arg Arg Leu Leu Ala  
 165 170 175  
 Leu Ile Tyr Leu Ala His Ala Ser Asp Val Leu Glu Asn Ala Phe Ala  
 180 185 190  
 Pro Leu Leu Asp Glu Gln Tyr Asp Leu Ala Thr Lys Arg Val Arg Gln  
 195 200 205  
 Leu Leu Asp Leu Asp Pro Glu Val Glu Cys Leu Lys Ala Asn Thr Asn  
 210 215 220  
 Glu Val Leu Trp Ala Val Val Ala Phe Thr Lys Xaa  
 225 230 235  
 Met Ala Gln Arg Met Val Trp Val Asp Leu Glu Met Thr Gly Leu Asp  
 1 5 10 15  
 Ile Glu Lys Asp Gln Ile Ile Glu Met Ala Cys Leu Ile Thr Asp Ser  
 20 25 30  
 Asp Leu Asn Ile Leu Ala Glu Gly Pro Asn Leu Ile Ile Lys Gln Pro  
 35 40 45  
 Asp Glu Leu Leu Asp Ser Met Ser Asp Trp Cys Lys Glu His His Gly  
 50 55 60  
 Lys Ser Gly Leu Thr Lys Ala Val Lys Glu Ser Thr Ile Thr Leu Gln  
 65 70 75 80  
 Gln Ala Glu Tyr Glu Phe Leu Ser Phe Val Arg Gln Gln Thr Pro Pro

(2) INFORMATION FOR SEQ ID NO: 488:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 200 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 488:

650

85 90 95  
 Gly Leu Cys Pro Leu Ala Gly Asn Ser Val His Glu Asp Lys Lys Phe  
 100 105 110  
 5 Leu Asp Lys Tyr Met Pro Gln Phe Met Lys His Leu His Tyr Arg Ile  
 115 120 125  
 Ile Asp Val Ser Thr Val Lys Glu Leu Cys Arg Arg Trp Tyr Pro Glu  
 130 135 140  
 10 Glu Tyr Glu Phe Ala Pro Lys Lys Ala Ser His Arg Ala Leu Asp  
 145 150 155 160  
 15 Asp Ile Ser Glu Ser Ile Lys Glu Leu Gln Phe Tyr Arg Asn Asn Ile  
 165 170 175  
 Phe Lys Lys Lys Ile Asp Glu Lys Lys Arg Lys Ile Ile Glu Asn Gly  
 180 185 190  
 20 Glu Asn Glu Lys Thr Val Ser Xaa  
 195 200

(2) INFORMATION FOR SEQ ID NO: 489:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 351 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 489:

Met Ala Thr Thr Ala Ala Pro Ala Gly Gly Ala Arg Asn Gly Ala Gly  
 1 5 10 15  
 Pro Glu Trp Gly Gly Phe Glu Glu Asn Ile Gln Gly Gly Gly Ser Ala  
 20 25 30  
 40 Val Ile Asp Met Glu Asn Met Asp Asp Thr Ser Gly Ser Ser Phe Glu  
 35 40 45  
 Asp Met Gly Glu Leu His Gln Arg Leu Arg Glu Glu Glu Val Asp Ala  
 50 55 60  
 45 Asp Ala Ala Asp Ala Ala Ala Glu Glu Glu Asp Gly Glu Phe Leu  
 65 70 75 80  
 Gly Met Lys Gly Phe Lys Gly Gln Leu Ser Arg Gln Val Ala Asp Gln  
 85 90 95  
 Met Trp Gln Ala Gly Lys Arg Gln Ala Ser Arg Ala Phe Ser Leu Tyr  
 100 105 110  
 55 Ala Asn Ile Asp Ile Leu Arg Pro Tyr Phe Asp Val Glu Pro Ala Gln  
 115 120 125  
 Val Arg Thr Gly Leu Leu Glu Ser Met Ile Pro Ile Lys Met Val Asn  
 130 135 140  
 60

651

Phe Pro Gln Lys Ile Ala Gly Glu Leu Tyr Gly Pro Leu Met Leu Val  
 145 150 155 160  
 Phe Thr Leu Val Ala Ile Leu Leu His Gly Met Lys Thr Ser Asp Thr  
 165 170 175  
 5 Ile Ile Arg Glu Gly Thr Leu Met Gly Thr Ala Ile Gly Thr Cys Phe  
 180 185 190  
 Gly Tyr Trp Leu Gly Val Ser Ser Phe Ile Tyr Phe Leu Ala Tyr Leu  
 195 200 205  
 10 Cys Asn Ala Gln Ile Thr Met Leu Gln Met Leu Ala Leu Leu Gly Tyr  
 210 215 220  
 Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr Tyr Asn Ile His  
 225 230 235 240  
 Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val Gly Gly Leu Ser  
 245 250 255  
 20 Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr Val Gly Pro Thr  
 260 265 270  
 Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu His Met Leu Phe  
 275 280 285  
 Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Val Gly Ile Leu  
 290 295 300  
 30 Asp Thr Leu Glu Gly Pro Asn Ile Pro Pro Ile Gln Arg Val Pro Arg  
 305 310 315 320  
 Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro Thr Thr Val Leu  
 325 330 335  
 Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln Ser His Xaa  
 340 345 350  
 40

(2) INFORMATION FOR SEQ ID NO: 490:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 265 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 490:

Met Arg Gly Ser Arg Gly Gly Trp Ala Gly Glu Met Ala Ala Ser Gly  
 1 5 10 15  
 Glu Ser Gly Thr Ser Gly Gly Gly Ser Thr Glu Glu Ala Phe Met  
 20 25 30  
 55 Thr Phe Tyr Ser Glu Val Lys Gln Ile Glu Lys Arg Asp Ser Val Leu  
 35 40 45  
 Thr Ser Lys Asn Gln Ile Glu Arg Leu Thr Arg Pro Gly Ser Ser Tyr  
 50 55 60

Phe Asn Leu Asn Pro Phe Glu Val Leu Gln Ile Asp Pro Glu Val Thr  
65 70 75 80

5 Asp Glu Glu Ile Lys Lys Arg Phe Arg Gln Leu Ser Ile Leu Val His  
85 90 95

Pro Asp Lys Asn Gln Asp Ala Asp Arg Ala Gln Lys Ala Phe Glu  
100 105 110

10 Ala Val Asp Lys Ala Tyr Lys Leu Leu Asp Gln Gln Lys Lys  
115 120 125

Arg Ala Leu Asp Val Ile Gln Ala Gly Lys Glu Tyr Val Glu His Thr  
130 135 140

Val Lys Glu Arg Lys Lys Gln Lys Lys Glu Gly Lys Pro Thr Ile  
145 150 155 160

20 Val Glu Glu Asp Asp Pro Glu Leu Phe Lys Gln Ala Val Tyr Lys Gln  
165 170 175

Thr Met Lys Leu Phe Ala Glu Leu Glu Ile Lys Arg Lys Glu Arg Glu  
180 185 190

25 Ala Lys Glu Met His Glu Arg Lys Arg Gln Arg Glu Glu Ile Glu  
195 200 205

Ala Gln Glu Lys Ala Lys Arg Glu Arg Glu Trp Gln Lys Asn Phe Glu  
210 215 220

Glu Ser Arg Asp Gly Arg Val Asp Ser Trp Arg Asn Phe Gln Ala Asn  
225 230 235 240

35 Thr Lys Gly Lys Lys Glu Lys Lys Asn Arg Thr Phe Leu Arg Pro Pro  
245 250 255

Lys Val Lys Met Glu Gln Arg Glu Xaa  
260 265

40

(2) INFORMATION FOR SEQ ID NO: 491:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 491:

50 Asp Ser Met Pro Thr Cys Pro Leu Xaa Ala Ser Leu Glu Cys Gly Pro  
1 5 10 15

Leu Leu Pro Val Arg Leu Cys Leu  
20 25

55

(2) INFORMATION FOR SEQ ID NO: 492:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 159 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 492:

5 Met Asn Glu Tyr Arg Val Pro Glu Leu Asn Val Gln Asn Gly Val Leu  
1 5 10 15

10 Lys Ser Leu Ser Phe Leu Phe Glu Tyr Ile Gly Glu Met Gly Lys Asp  
20 25 30

Tyr Ile Tyr Ala Val Thr Pro Leu Leu Glu Asp Ala Leu Met Asp Arg  
35 40 45

15 Asp Leu Val His Arg Gln Thr Ala Ser Ala Val Val Gln His Met Ser  
50 55 60

Leu Gly Val Tyr Gly Phe Gly Cys Glu Asp Ser Leu Asn His Leu Leu  
65 70 75 80

Asn Tyr Val Trp Pro Asn Val Phe Glu Thr Ser Pro His Val Ile Gln  
85 90 95

25 Ala Val Met Gly Ala Leu Glu Gly Leu Arg Val Ala Ile Gly Pro Cys  
100 105 110

Arg Met Leu Gln Tyr Cys Leu Gln Gly Leu Phe His Pro Ala Arg Lys  
115 120 125

30 Val Arg Asp Val Tyr Trp Lys Ile Tyr Asn Ser Ile Tyr Ile Gly Ser  
130 135 140

Gln Asp Ala Leu Ile Ala His Tyr Pro Arg Ile Tyr Gln Arg Xaa  
145 150 155

40

(2) INFORMATION FOR SEQ ID NO: 493:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 279 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 493:

45 Met Ile Ser Asp Asn Ser Ala Glu Asn Ile Ala Leu Val Thr Ser Met  
1 5 10 15

50 Tyr Asp Gly Leu Leu Gln Ala Gly Ala Arg Leu Cys Pro Thr Val Gln  
20 25 30

Leu Glu Asp Ile Arg Asn Leu Gln Asp Leu Thr Pro Leu Lys Leu Ala  
35 40 45

55 Ala Lys Glu Gly Lys Ile Glu Ile Phe Arg His Ile Leu Gln Arg Glu  
50 55 60

Phe Ser Gly Leu Ser His Leu Ser Arg Lys Phe Thr Glu Trp Cys Tyr  
65 70 75 80

Gly Pro Val Arg Val Ser Leu Tyr Asp Leu Ala Ser Val Asp Ser Cys  
85 90 95

5 Glu Glu Asn Ser Val Leu Glu Ile Ile Ala Phe His Cys Lys Ser Pro  
100 105 110

His Arg His Arg Met Val Val Leu Glu Pro Leu Asn Lys Leu Glu  
115 120 125

10 Ala Lys Trp Asp Leu Leu Ile Pro Lys Phe Phe Leu Asn Phe Leu Cys  
130 135 140

Asn Leu Ile Tyr Met Phe Ile Phe Thr Ala Val Ala Tyr His Gln Pro  
145 150 155 160

Thr Leu Lys Lys Gln Ala Ala Pro His Leu Lys Ala Glu Val Gly Asn  
165 170 175

20 Ser Met Leu Leu Thr Gly His Ile Leu Ile Leu Leu Gly Ile Tyr  
180 185 190

Leu Leu Val Gly Gln Leu Trp Tyr Phe Trp Arg Arg His Val Phe Ile  
195 200 205

25 Trp Ile Ser Phe Ile Asp Ser Tyr Phe Glu Ile Leu Phe Leu Gln  
210 215 220

Ala Leu Leu Thr Val Val Ser Gln Val Leu Cys Phe Leu Xaa Ile Glu  
225 230 235 240

Trp Tyr Leu Pro Leu Leu Val Ser Ala Leu Val Leu Gly Trp Leu Asn  
245 250 255

35 Leu Leu Tyr Tyr Thr Arg Gly Phe Gln His Thr Gly Ile Tyr Ser Val  
260 265 270

Met Ile Gln Lys Pro Trp Xaa  
275

40

(2) INFORMATION FOR SEQ ID NO: 494:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 193 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 494:

Met Ile Arg Cys Gly Leu Ala Cys Glu Arg Cys Arg Trp Ile Leu Pro  
1 5 10 15

Leu Leu Leu Leu Ser Ala Ile Ala Phe Asp Ile Ile Ala Leu Ala Gly  
20 25 30

Arg Gly Trp Leu Gln Ser Ser Asp His Gly Gln Thr Ser Ser Leu Trp  
35 40 45

60 Trp Lys Cys Ser Gln Glu Gly Gly Gly Ser Tyr Glu Glu Gly

50 55 60

Cys Gln Ser Leu Met Glu Tyr Ala Trp Gly Arg Ala Ala Ala Met  
65 70 75 80

Leu Phe Cys Gly Phe Ile Ile Leu Val Ile Cys Phe Ile Leu Ser Phe  
85 90 95

10 Phe Ala Leu Cys Gly Pro Gln Met Leu Val Phe Leu Arg Val Ile Gly  
100 105 110

Gly Leu Leu Ala Leu Ala Val Phe Gln Ile Ile Ser Leu Val Ile  
115 120 125

15 Tyr Pro Val Lys Tyr Thr Gln Thr Phe Thr Leu His Ala Asn Xaa Ala  
130 135 140

Val Thr Tyr Ile Tyr Asn Trp Ala Tyr Gly Phe Gly Trp Ala Thr  
145 150 155 160

20 Ile Ile Leu Ile Gly Cys Ala Phe Phe Cys Cys Leu Pro Asn Tyr  
165 170 175

Glu Asp Asp Leu Leu Gly Asn Ala Lys Pro Arg Tyr Phe Thr Ser  
180 185 190

Ala

30

(2) INFORMATION FOR SEQ ID NO: 495:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 205 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 495:

Met Ala Ala Gly Asp Gln Val Phe Ser Gly Ala Gly His Val Xaa Glu  
1 5 10 15

His Val Ala Gly Arg His Ala Trp Leu Leu Thr Trp Gln Ser Ala  
20 25 30

Cys Pro Ala Asn Arg Leu Ser Leu Val Pro Leu Val Pro Ser Ala Ser  
35 40 45

Met Thr Arg Leu Met Arg Xaa Arg Thr Ala Ser Gly Ser Ser Val Ile  
50 55 60

Leu Trp Met Ala Pro Ala Ala Pro Thr Pro Ala Arg Ala Pro Glu  
65 70 75 80

55 Ala Ala Pro Thr Pro Ala Arg Ala Pro Ala Ala Arg Thr Pro Ala  
85 90 95

Arg Gly Pro Thr Trp Thr Ser Pro Thr Arg Val Leu Leu Gly Thr  
100 105 110



Xaa Pro Gly Pro Ser Pro Trp Arg Ser Pro Ala Arg Arg Pro Ala Gln  
115 120 125

5 Leu Pro Pro Asp Ser Asp Leu Cys Ser Gly Pro Leu Leu Pro Gly  
130 135 140

Pro Phe Ser Pro Pro Ala Cys His Thr Ala Pro Ann Ser Val Leu Ile  
145 150 155 160

10 Gln Ser Leu Phe Cys Lys Ser Glu Leu Trp Trp Arg Gln Met Arg Ser  
165 170 175

Ile Thr Trp Val Pro Ser Pro Lys Ala Gly Trp Arg Trp Thr Lys Gly  
180 185 190

15 Arg Lys Gln Ala Ser Pro His Arg Ile Leu Phe His Xaa  
195 200 205

20 (2) INFORMATION FOR SEQ ID NO: 496:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 147 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 496:

30 Met Ala Leu Thr Leu Leu Pro Ser Val Ser Arg Leu Pro Gly Glu Arg  
1 5 10 15

Met Ala Ala Ser Gly Leu Pro Tyr Val Leu His His Lys Ser Ser Leu  
20 25 30

35 Met Lys Val Ile Phe Phe Pro Tyr Pro Val Leu Pro Leu Pro Ala Pro  
35 40 45

Asn Gly Thr Trp Val Pro Arg Leu Val Leu Gly Leu Gly Ser Gly Asp  
50 55 60

40 Gln Val His Tyr Leu Pro Ile Ser Ser Ile Val Asn Tyr Gly Thr  
65 70 75 80

Ser Val Ser Gly Lys Ser Trp Val Phe Leu Val Tyr Pro Leu His Pro  
85 90 95

Thr Pro Thr Trp Ser Thr Arg Cys Phe Gln Val Trp Asp Leu Leu Ser  
100 105 110

50 Val Glu Leu Pro Asp Lys Gly Glu Gly Asn Thr Arg Arg Ala Ser Gly  
115 120 125

Val Pro Gly Leu Ser Gln Leu Pro Thr Ser His Lys Pro Ile Lys Gln  
130 135 140

Glu Tyr Xaa

145

60

(2) INFORMATION FOR SEQ ID NO: 497:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 64 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 497:

5 Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val  
1 5 10 15

10 Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys Glu Val Gly  
20 25 30

15 Ala Ala Leu Pro Pro Arg Gly Pro Ser Leu Ser Asp Cys Leu Gly Leu  
35 40 45

Pro Pro Trp Thr Pro Trp Gly Pro Ala Trp Thr Leu Ala Gln Ser Xaa  
50 55 60

25 (2) INFORMATION FOR SEQ ID NO: 498:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 94 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 498:

35 Met Ser Thr Gly Ala Leu Ann Thr Ser Pro Pro Ala Ser Ann Arg Leu  
1 5 10 15

Glu Ser Thr Leu Ann Glu Tyr Leu Ile Gln Pro Gln Leu His Cys Ser  
20 25 30

40 Ser Val Gln Arg Leu Thr Leu Lys Trp Gly Cys Ser Ser Leu Gln Arg  
35 40 45

Asp Gly Gln Ala Val Pro Trp Gly Leu Trp Gln Arg Ala Tyr Pro Ser  
50 55 60

Leu Leu Pro Thr Leu Pro Ser Asp Leu Leu Arg Pro His Ala Val Thr  
65 70 75 80

Pro Ser Val Ser Val Ser Val His Thr Cys Glu Ser Ser Xaa  
85 90

55 (2) INFORMATION FOR SEQ ID NO: 499:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 499:

Met Phe Leu Ile Phe Val Tyr Phe Leu Lys Xaa Leu Phe Ser Ser Ser  
1 5 10 15

5 Leu Pro Phe Leu Trp Leu  
20

10 (2) INFORMATION FOR SEQ ID NO: 500:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 500:

Arg Gly Gly Leu Cys Pro Leu Leu Val Pro Gly Pro Leu Ala Arg Gln  
1 5 10 15

20 Glu Pro Ser Pro Ser Leu Gln Gly Cys Ser Glu Ser Pro Val Gly Met  
25

Asp

30 (2) INFORMATION FOR SEQ ID NO: 501:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 501:

Met Gln Phe Leu Leu Thr Ala Phe Leu Leu Val Pro Leu Leu Ala Leu  
1 5 10 15

40 Cys Asp Val Pro Ile Ser Leu Gly Phe Ser Pro Ser  
20 25

45 (2) INFORMATION FOR SEQ ID NO: 502:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 15 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 502:

Pro Gly Lys Pro Gln Ala Cys Pro Glu Leu Thr Ser Val Leu Pro  
1 5 10 15

55 (2) INFORMATION FOR SEQ ID NO: 503:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

60

- (A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 503:

5 Asn Lys Ser Leu Xaa Ser Cys Leu Phe Val Leu His Phe Val Leu His  
1 5 10 15

10 Cys Xaa Phe

15 (2) INFORMATION FOR SEQ ID NO: 504:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 504:

20 Met Glu Lys Thr His Arg Leu Arg Ile Arg Asn Pro Cys Leu Gln Phe  
1 5 10 15

25 Ser Ile Leu Asn Leu Phe Leu Leu Lys Met Ile Val Ser  
20 25

30 (2) INFORMATION FOR SEQ ID NO: 505:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 505:

35 Met Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln  
1 5 10 15

40 Asn Leu Ser Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu  
20 25 30

45 Ala Gly Lys Leu Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly  
35 40 45

Pro Ser Ser Phe Gln Asn Pro Ala Ser Ser Pro Ser Ser Trp Thr His  
50 55 60

50 Glu Glu Glu Pro Gly Tyr Phe Pro Gln Tyr Xaa  
65 70 75

55 (2) INFORMATION FOR SEQ ID NO: 506:

(1) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 10 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

60

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 506:

Leu Pro Leu Ala Glu Leu Lys Asn Trp Val  
1 5 10

5

(2) INFORMATION FOR SEQ ID NO: 507:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 207 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 507:

Met Leu Trp Phe Gly Gly Cys Ser Ala Val Asn Ala Thr Gly His Leu  
1 5 10 15

Ser Asp Thr Leu Trp Leu Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr  
20 25 30

Gly Asp Val Val Pro Gly Thr Met Trp Gly Lys Ile Val Cys Leu Cys  
35 40 45

Thr Gly Val Met Gly Val Cys Cys Thr Ala Leu Leu Val Ala Val  
50 55 60

Ala Arg Lys Leu Glu Phe Asn Lys Ala Glu Lys His Val His Asn Phe  
65 70 75 80

Met Met Asp Ile Gln Tyr Thr Lys Glu Met Lys Glu Ser Ala Ala Arg  
85 90 95

Val Leu Gln Glu Ala Trp Met Phe Tyr Lys His Thr Arg Arg Lys Glu  
100 105 110

Ser His Ala Ala Arg Arg His Gln Arg Xaa Leu Leu Ala Ala Ile Asn  
115 120 125

Ala Phe Arg Gln Val Arg Leu Lys His Arg Lys Leu Arg Glu Gln Val  
130 135 140

Asn Ser Met Val Asp Ile Ser Lys Met His Met Ile Leu Tyr Asp Leu  
145 150 155 160

Gln Gln Asn Leu Ser Ser His Arg Ala Leu Glu Lys Gln Ile Asp  
165 170 175

Thr Leu Ala Gly Lys Leu Asp Ala Leu Thr Gly Leu Leu Ser Thr Ala  
180 185 190

Leu Gly Pro Arg Gln Leu Pro Glu Pro Ser Gln Gln Ser Lys Xaa  
195 200 205

55

(2) INFORMATION FOR SEQ ID NO: 508:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 508:

Met Trp Arg Cys Arg Gly Lys Leu Ser Phe Pro Leu Phe Ala Val Val  
1 5 10 15

Ile Val Ser Cys Arg Lys Asp Gly Pro Asp Ala Ala Ala Pro Ala  
20 25 30

Val Xaa Lys Lys  
35

10

(2) INFORMATION FOR SEQ ID NO: 509:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 509:

Met Ala Leu Val Ala Leu Phe Thr Gln Leu Met Arg Xaa Leu Gly Arg  
1 5 10 15

Cys Pro Gln

30

(2) INFORMATION FOR SEQ ID NO: 510:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 32 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 510:

Met Thr Phe Pro Phe Glu Lys Glu Asn Ser Cys Phe Gln Cys Leu Leu  
1 5 10 15

Phe Asp Ser Trp Arg Glu Gln Thr Arg Thr Asn Ile Gln Pro Gln Arg  
20 25 30

45

(2) INFORMATION FOR SEQ ID NO: 511:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 28 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 511:

Met His Leu Leu Asp Phe Phe Arg Asp Leu Val Leu Leu Val Leu Leu  
1 5 10 15

60

662

663

Ala Leu Leu Asp Ser Phe Trp Leu Glu Val Gln Lys  
20 25

5

(2) INFORMATION FOR SEQ ID NO: 512:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 512:

10

Met Cys Leu Ile His Phe Ile Lys Ile Ile Leu Val Phe Ile Leu Lys  
1 5 10 15

Leu Trp Leu Tyr Ser Gln Lys Cys Pro Lys  
20 25

20

(2) INFORMATION FOR SEQ ID NO: 513:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 33 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 513:

25

Met Ile His Val His Glu Trp Asn Asp Gln Met Leu Met Val Tyr Ile  
1 5 10 15

Phe Leu Tyr Pro Val Ser Ile Thr Phe Leu Asn Leu Cys Ser Leu Thr  
20 25 30

Cys

35

(2) INFORMATION FOR SEQ ID NO: 514:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 514:

45

Leu Asn Glu Ser Tyr Val Ser Arg Ala Gly Gly Trp Phe Ser Met Phe  
1 5 10 15

Xaa Leu Ile Phe Phe Leu Leu Ala Leu Gly Ser Xaa Leu Cys Leu Leu  
20 25 30

55

Leu Cys Leu Pro Ser Phe Asn Lys Thr Arg Lys Gln Lys Pro  
35 40 45

60

(2) INFORMATION FOR SEQ ID NO: 515:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 515:

5

Ser Ser Lys Thr Pro Leu Pro Ser Glu Arg Arg Trp Ile Ser Gly Ser  
1 5 10 15

Ser Leu Met Ala Pro Arg Pro Trp Leu Leu Gly Ile Ala Leu Leu Gly  
20 25 30

Leu Trp Ala Leu Glu Pro Ala Leu Gly His Trp  
35 40

(2) INFORMATION FOR SEQ ID NO: 516:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 3 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 516:

25

Leu Asn Trp  
1

30

(2) INFORMATION FOR SEQ ID NO: 517:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 174 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 517:

35

Phe Ala Phe Cys Ala Glu Leu Met Ile Gln Asn Trp Thr Leu Gly Ala  
1 5 10 15

Val Asp Ser Gln Met Asp Asp Met Asp Met Asp Leu Asp Lys Glu Phe  
20 25 30

Leu Gln Asp Leu Lys Glu Leu Lys Val Leu Val Ala Asp Lys Asp Leu  
35 40 45

Leu Asp Leu His Lys Ser Leu Val Cys Thr Ala Leu Arg Gly Lys Leu  
50 55 60

Gly Val Phe Ser Glu Met Glu Ala Asn Phe Lys Asn Leu Ser Arg Gly  
65 70 75 80

Leu Val Asn Val Ala Ala Lys Leu Thr His Asn Lys Asp Val Arg Asp  
85 90 95

Leu Phe Val Asp Leu Val Glu Lys Phe Val Glu Pro Cys Arg Ser Asp  
100 105 110

His Trp Pro Leu Ser Asp Val Arg Phe Leu Asn Gln Tyr Ser Ala  
115 120 125

5 Ser Val His Ser Leu Asp Gly Phe Arg His Gln Ala Ser Gly Thr Ala  
130 135 140

Thr Trp Ala Pro Ser Ala Ala Ser Cys Ala Cys Ile Met Thr Glu  
145 150 155 160

10 Val Pro Pro Asn Ala Pro Thr Leu Thr Ile Lys Leu Leu  
165 170

15 (2) INFORMATION FOR SEQ ID NO: 518:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 518:

Met Trp Lys Asn Leu Gly Ser Gly Ser Val Phe Val Thr Trp Phe Ser  
1 5 10 15

Leu Val Met Ile Leu Ser Gly Ile Gly Pro Leu Gly Asp Ala Glu Asp  
20 25 30

30 Ser Ile Ser Asp Val Ser His Arg Leu Arg Pro  
35 40

35 (2) INFORMATION FOR SEQ ID NO: 519:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 13 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 519:

Phe Gln Phe Pro Leu Leu Thr Ile Ala Leu Gln Phe Leu  
1 5 10

(2) INFORMATION FOR SEQ ID NO: 520:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 520:

Met His Tyr Val Ile Val Leu Ser Leu Phe Val Val Leu Glu Lys Lys  
1 5 10 15

Asn Lys Met Gly Ser Asp Gly Cys Leu Arg Lys Asn Gly Ser  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 521:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 47 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 521:

Met Ser Arg Ser Ile Val Leu Arg Gly Ser Leu Phe Leu Phe Ser  
1 5 10 15

His Tyr Thr Leu Lys Leu Ser Val Ile Lys Gln Thr Asn Arg Lys  
20 25 30

Ile Val Trp Glu Lys Pro Cys Ile Arg Leu Phe Tyr Xaa Val Leu  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 522:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 26 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 522:

Met Pro Leu Pro Val Leu Leu Cys Leu Thr Leu Pro Met Pro Leu Pro  
1 5 10 15

Ser Ala Thr Ala Arg Gly Gly Asn Arg Thr  
20 25

(2) INFORMATION FOR SEQ ID NO: 523:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 58 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 523:

Ser Ser Ile Pro Val Ser Ile Leu Ile Gly Met Lys Leu Ile Leu Tyr  
1 5 10 15

Leu Leu Ile Thr Glu Ser Gly Ser His Glu Lys Lys Ser Phe Tyr Pro  
20 25 30

Ser Phe Lys Tyr Met Phe Lys Ile Ile Ile Tyr Val Ser Ala Tyr Cys  
35 40 45

Arg Thr Ala Leu Arg Ala Thr Val Ser His  
50 55

60

666

667

## (2) INFORMATION FOR SEQ ID NO: 524:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 524:

5 Asn Arg Thr Leu Leu Phe Leu Ile Leu Phe Val Leu Phe Gly Leu Gly  
1 5 10 15

Tyr Gly Phe

15

## (2) INFORMATION FOR SEQ ID NO: 525:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 40 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 525:

25 Met Phe Leu Leu Val Leu Ser Val Phe Cys Asp Phe Met Cys Ser Ile  
1 5 10 15

Ala Pro Arg Cys His Ala Leu Ser Leu Val Ser Leu Arg Ala Gln His  
20 25 30

Leu Ser Leu Phe Ile Thr Cys His  
35 40

30

35

## (2) INFORMATION FOR SEQ ID NO: 526:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 526:

45 Met Leu Leu Phe Ile Leu Leu Thr Leu Ser Ser Gly Cys Arg Leu Leu  
1 5 10 15

Val Ser Ser Trp Lys Thr Phe Leu Pro His Phe Ser Leu Pro Gly Pro  
20 25 30

50 Arg Glu His Pro Glu Gly Ser Arg Thr Trp Phe Arg Tyr Trp Glu  
35 40 45

Pro Gly Ala His Cys Leu His Cys Ala  
50 55

55

## (2) INFORMATION FOR SEQ ID NO: 527:

- (1) SEQUENCE CHARACTERISTICS:

60 Ser Pro Gly Leu Lys Asn Gly Ile Phe Leu Phe Leu Leu Arg Ala Met  
100 105 110

WO 98/39448

- (A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 527:

5 Ala Arg Leu Leu Leu Phe Leu Ser Ser Val His Pro Ser Ile Met Pro  
1 5 10 15

Ser Cys Asn Gln Leu  
20

10

## (2) INFORMATION FOR SEQ ID NO: 528:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 528:

20 Met Ser Leu Thr Ser Ser Leu Thr Phe Leu Ser His Ile Leu Leu Leu  
1 5 10 15

25 Pro Gln Lys Leu Gln Phe Leu Ser Trp Met Glu Arg Gln Gln Arg Cys  
20 25 30

Thr Gly Val Ala Lys Tyr Ala  
35

30

## (2) INFORMATION FOR SEQ ID NO: 529:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 128 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 529:

40 Met Val Leu Arg Leu Ile Gln Leu Ile Phe Leu Ile Phe Ile His  
1 5 10 15

45 Ile Ile Ile Leu Leu Ile Pro Gly Ser Arg Pro Cys Gly Ser Trp Val  
20 25 30

Asn Asp Arg Xaa Leu Gly Leu Arg Asp Val Thr His Leu Ile Tyr Leu  
35 40 45

50 His Trp Val His Gly His Leu Pro Trp Cys His Pro Tyr Ile Gln Val  
50 55 60

Glu Phe Ser Ala Leu Ile Glu Ser Thr Ala Gln Leu Gly Leu Pro Phe  
65 70 75 80

55 Ser Trp Val Arg Val Ile His Pro Phe Leu Val Leu Pro Cys Leu Tyr  
85 90 95

60 Ser Pro Gly Leu Lys Asn Gly Ile Phe Leu Phe Leu Leu Arg Ala Met  
100 105 110

Pro Gly Gly Met Phe Pro Gly Asn Leu Glu Ala Phe Arg Val Pro Val  
115 120 125

5

10 (2) INFORMATION FOR SEQ ID NO: 530:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 82 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 530:

Met Gly Ser Ser Val Leu Pro Phe Cys Val Cys Val Thr Ser Pro Ser  
1 5 10 15

20 Leu Gly Gly Arg Cys Ile Gln Gly Arg Phe Ala Ser His Ser Lys Phe  
20 25 3025 Trp Gly Phe Gly Xaa Lys Thr Ala Ser Phe Gly Ala Val Gly Glu Thr  
35 40 45

Pro Pro Asp Gln Glu Pro Gln Lys Thr Glu Thr Glu Pro Ala Thr Ser Ser  
50 55 60

30 His Ala Arg Pro Trp Ala Arg Val Ile Gly Leu Arg Ile Trp Pro Gln  
65 70 75 80

Pro Asn

35

(2) INFORMATION FOR SEQ ID NO: 531:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 20 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 531:

Met Leu Leu Ser Val Ala Ile Phe Ile Leu Leu Thr Leu Val Tyr Ala  
1 5 10 15

50 Tyr Trp Thr Met  
20

(2) INFORMATION FOR SEQ ID NO: 532:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 75 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 532:

55

60

Asn Cys Glu Ile Leu Glu Tyr Cys Tyr Tyr Leu Thr Gln Leu Lys Ile  
1 5 10 15

5 Ser Met Gly Lys Tyr Leu Ser Ile Pro Thr Val Leu Leu Lys Ile Ile  
20 25 3010 Arg Cys Ser Ile Thr Ala Val Ser Asp Ser Ser Thr Ser Trp Ala Ile  
35 40 45

Lys Ala Gln Leu Lys Ile Glu Asn Lys Asp Leu Asn Lys Thr Ala  
50 55 60

15 Lys Gly Gly Gly Gln Glu Ala Lys Thr Cys Thr  
65 70 75

(2) INFORMATION FOR SEQ ID NO: 533:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 60 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 533:

Met Phe Leu Met Arg Met His Leu Cys Phe Cys Lys Tyr Cys Cys Ser  
1 5 10 15

30 Phe Ile Val Thr Pro Thr Ser Thr Ser Asn Thr Xaa Ser Tyr Leu Trp  
20 25 3035 Pro Trp Ile Ser Ala Ser Met Ala Gly Arg Gly Ser Xaa Trp Ala Cys  
35 40 45

Thr Leu Asn Ala Val Thr Arg Glu Gly Leu Pro Glu  
50 55 60

40 (2) INFORMATION FOR SEQ ID NO: 534:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 39 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 534:

50 Met Ser Leu Leu Asn Thr His Thr Leu Cys Phe Val Leu Phe Cys Phe  
1 5 10 15

Thr Leu Ser Ile Asn Gln Glu Lys Leu Ala Asn His Leu Ala Phe Arg  
20 25 30

55 Ile Leu Phe Phe Ile Val Phe  
35

60 (2) INFORMATION FOR SEQ ID NO: 535:

670

671

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 2 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 535:

5

Met Leu

1

10

- (2) INFORMATION FOR SEQ ID NO: 536:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 536:

15

Met Asp Gln Phe Lys Ile Phe Tyr Phe Leu Lys Ala Phe Phe Ala Cys

1

5

10

15

Cys Asn Val Gln Asp Pro Ser Pro Phe Met Gly Glu Thr Gly Ser Tyr

1

20

25

30

Leu Asn Ile Gly

35

30

- (2) INFORMATION FOR SEQ ID NO: 537:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 14 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 537:

35

Met Phe Asp Phe Leu Ser Tyr Phe Lys Asp Leu Leu Ser Cys

1

5

10

- (2) INFORMATION FOR SEQ ID NO: 538:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 538:

50

Met Gly Phe Gly Phe Val Leu Asn Ile Phe Ser Phe Phe Leu Xaa Pro

1

5

10

15

Pro Leu

1

5

10

15

20

25

30

35

40

45

50

55

60

- (2) INFORMATION FOR SEQ ID NO: 539:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 11 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 539:

5

Leu Leu Leu Trp Thr Leu Leu Ala Xaa Tyr Xaa

1

5

10

- (2) INFORMATION FOR SEQ ID NO: 540:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 108 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 540:

15

Met Ala Ala Gln Lys Asp Gln Gln Lys Asp Ala Glu Ala Glu Gly Leu

1

5

10

15

Ser Gly Thr Thr Leu Leu Pro Lys Leu Ile Pro Ser Gly Ala Gly Arg

1

20

25

30

Glu Trp Leu Glu Arg Arg Ala Thr Ile Arg Pro Trp Ser Thr Phe

1

35

40

45

Val Asp Gln Gln Arg Phe Ser Arg Pro Arg Asn Leu Gly Glu Leu Cys

1

50

55

60

Gln Arg Leu Val Arg Asn Val Glu Tyr Tyr Gln Ser Asn Tyr Val Phe

1

65

70

75

80

Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met Leu

1

85

90

95

Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Xaa

1

100

105

- (2) INFORMATION FOR SEQ ID NO: 541:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 106 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 541:

45

Phe Val Phe Leu Gly Leu Ile Leu Tyr Cys Val Val Thr Ser Pro Met

1

5

10

15

Leu Leu Val Ala Leu Ala Val Phe Phe Gly Ala Cys Tyr Ile Leu Tyr

1

20

25

30

Leu Arg Thr Leu Glu Ser Lys Leu Val Leu Phe Gly Arg Glu Val Ser

1

35

40

45



Pro Ala His Gln Tyr Ala Leu Ala Gly Ile Ser Phe Pro Phe  
50 55 60

5 Trp Leu Ala Gly Ala Gly Ser Ala Val Phe Trp Val Leu Gly Ala Thr  
65 70 75 80

Leu Val Val Ile Gly Ser His Ala Phe His Gln Ile Glu Ala Val  
85 90 95

10 Asp Gly Glu Glu Leu Gln Met Glu Pro Val  
100 105

15 (2) INFORMATION FOR SEQ ID NO: 542:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 136 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 542:

Met Asp Arg Phe Thr Val Ala Gly Val Leu Pro Asp Ile Glu Gln Phe  
1 5 10 15

Phe Asn Ile Gly Asp Ser Ser Gly Leu Ile Gln Thr Val Phe Ile  
20 25 30

Ser Ser Tyr Met Val Leu Ala Pro Val Phe Gly Tyr Leu Gly Asp Arg  
35 40 45

Tyr Asn Arg Lys Tyr Leu Met Cys Gly Gly Ile Ala Phe Thr Ser Leu  
50 55 60

Val Thr Leu Gly Ser Ser Phe Ile Pro Gly Glu His Phe Thr Leu Leu  
65 70 75 80

Leu Leu Thr Arg Gly Leu Val Gly Val Gly Glu Ala Ser Tyr Ser Thr  
85 90 95

Ile Ala Pro Thr Leu Ile Ala Asp Leu Phe Val Ala Asp Gln Arg Thr  
100 105 110

Gly Cys Ser Ala Ser Thr Leu Pro Phe Arg Trp Ala Val Val Trp  
115 120 125

Ala Thr Leu Gln Ala Pro Lys Xaa  
130 135

(2) INFORMATION FOR SEQ ID NO: 543:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 424 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 543:

Met Ala Gly Asp Trp His Trp Ala Leu Arg Val Thr Pro Gly Leu Gly  
1 5 10 15

5 Val Val Ala Val Leu Leu Phe Leu Val Val Arg Glu Pro Pro Arg  
20 25 30

Gly Ala Val Glu Arg His Ser Asp Leu Pro Pro Leu Asn Pro Thr Ser  
35 40 45

10 Trp Trp Ala Asp Leu Arg Ala Leu Ala Arg Asn Pro Ser Phe Val Leu  
50 55 60

Ser Ser Leu Gly Phe Thr Ala Val Ala Phe Val Thr Gly Ser Leu Ala  
65 70 75 80

15 Leu Trp Ala Pro Ala Phe Leu Leu Arg Ser Arg Val Val Leu Gly Glu  
85 90 95

20 Thr Pro Pro Cys Leu Pro Gly Asp Ser Cys Ser Ser Ser Asp Ser Leu  
100 105 110

Ile Phe Gly Leu Ile Thr Cys Leu Thr Gly Val Leu Gly Val Gly Leu  
115 120 125

25 Gly Val Glu Ile Ser Arg Arg Xaa Arg His Ser Asn Pro Arg Ala Asp  
130 135 140

30 Pro Leu Val Cys Ala Thr Gly Leu Leu Gly Ser Ala Pro Phe Leu Phe  
145 150 155 160

Leu Ser Leu Ala Cys Ala Arg Gly Ser Ile Val Ala Thr Tyr Ile Phe  
165 170 175

35 Ile Phe Ile Gly Glu Thr Leu Leu Ser Met Asn Trp Ala Ile Val Ala  
180 185 190

Asp Ile Leu Leu Tyr Val Val Ile Pro Thr Arg Arg Ser Thr Ala Glu  
195 200 205

40 Ala Phe Gln Ile Val Leu Ser His Leu Leu Gly Asp Ala Gly Ser Pro  
210 215 220

Tyr Leu Ile Gly Leu Ile Ser Asp Arg Leu Arg Asn Trp Pro Pro  
225 230 235 240

Ser Phe Leu Ser Glu Phe Arg Ala Leu Gln Phe Ser Leu Met Leu Cys  
245 250 255

50 Ala Phe Val Gly Ala Leu Gly Gly Ala Leu Ser Trp Ala Pro Xaa Ser  
260 265 270

Ser Leu Arg Pro Thr Ala Gly Gly His Ser Cys Thr Cys Arg Ala Cys  
275 280 285

55 Cys Thr Lys Gln Gly Pro Gln Thr Thr Gly Leu Trp Cys Pro Ser Gly  
290 295 300

Ala Ala Pro Ala Cys Pro Trp Pro Val Cys Ser Ser Glu Arg Leu  
305 310 315 320

60

674

Pro Leu Thr Tyr Leu His Ile Cys His Ser Xaa Pro Trp Ala His Pro 335  
330  
Thr Lys Gly Leu Gly Leu Thr Pro Trp Pro Gly Pro Ala Ser Arg Gly 350  
345  
Thr Leu Gly Arg Val Pro Ala Pro Arg His Tyr Xaa Gly Ser Ser Gly 365  
355  
Glu Glu Val Gly Val Gln Glu Gly Asp Pro Ser Pro Gln Gly Xaa Pro 380  
370  
Gln Gly Leu Gly Ala Ile Cys Asn Gly Ile Lys Phe Val Ala Arg Pro 400  
385  
Gln Val Pro Ala Leu Val Phe Leu Trp Val Ala Ser Asp Leu Ala Pro 415  
405  
Arg Leu His Pro Arg Ala Pro Glu 420

5

10

15

20

(2) INFORMATION FOR SEQ ID NO: 544:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 544:

Met Phe Arg Phe Val Ile Cys Leu Phe Leu Trp Leu Val Leu Cys Arg 15  
1  
Asp Ser Thr Ser Ala Ser Arg Ile Ala Leu Tyr Tyr Arg Ile Val Phe 30  
20  
Leu Ile His Gln Cys Ser Ser 35

35

40

(2) INFORMATION FOR SEQ ID NO: 545:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 545:

Met Leu Pro Trp Xaa Ala Gln Leu Leu Asp Arg Thr Ile Gly Pro Leu 15  
1  
Tyr Leu Leu Phe Val Gln Phe Ser Pro Ala Phe Ser Arg Thr Ser Pro 30  
20  
Trp Arg Ser Pro Lys Asn Phe Arg Arg Leu Tyr Pro Pro Cys Thr Thr 45  
35  
Ser Gly Cys Ala Ala Arg Trp Leu Phe Ser 40

55

60

675

50 55  
(2) INFORMATION FOR SEQ ID NO: 546:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 33 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 546:

Met Gly Leu Ser Val Leu Leu Pro Leu Cys Leu Leu Gly Pro Gly Arg 15  
1  
Phe Thr Ser Gly Gln Lys Pro Leu Asp Thr Pro Gly Leu Gly Val Pro 30  
20  
Phe

20

(2) INFORMATION FOR SEQ ID NO: 547:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 367 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 547:

Met Ala Lys Pro Gln Val Val Ala Pro Val Leu Met Ser Lys Leu 15  
1  
Ser Val Asn Ala Pro Glu Phe Tyr Pro Ser Gly Tyr Ser Ser Ser Tyr 25  
20  
Thr Glu Ser Tyr Glu Asp Gly Cys Glu Asp Tyr Pro Thr Leu Ser Glu 45  
35  
Tyr Val Gln Asp Phe Leu Asn His Leu Thr Glu Gln Pro Gly Ser Phe 55  
50  
Glu Thr Glu Ile Glu Gln Phe Ala Glu Thr Leu Asn Gly Cys Val Thr 80  
65  
Thr Asp Asp Ala Leu Gln Glu Leu Val Glu Leu Ile Tyr Gln Gln Ala 95  
85  
Thr Ser Ile Pro Asn Phe Ser Tyr Met Gly Ala Arg Leu Cys Asn Tyr 110  
100  
Leu Ser His His Leu Thr Ile Ser Pro Gln Ser Gly Asn Phe Arg Gln 125  
115  
Leu Leu Leu Gln Arg Cys Arg Thr Glu Tyr Glu Val Lys Asp Gln Ala 140  
130  
Ala Lys Gly Asp Glu Val Thr Arg Lys Arg Phe His Ala Phe Val Leu 160  
145  
150  
155

25

30

35

40

45

50

55

60

Phe Leu Gly Glu Leu Tyr Leu Asn Leu Glu Ile Lys Gly Thr Asn Gly 165 170 175  
 5 Gln Val Thr Arg Ala Asp Ile Leu Gln Val Gly Leu Arg Glu Leu Leu 180 185 190  
 Asn Ala Leu Phe Ser Asn Pro Met Asp Asp Asn Leu Ile Cys Ala Val 195 200 205  
 10 Lys Leu Leu Lys Leu Thr Gly Ser Val Leu Glu Asp Ala Trp Lys Glu 210 215 220  
 Lys Gly Lys Met Asp Met Glu Glu Ile Ile Gln Arg Ile Glu Asn Val 225 230 235 240  
 15 Val Leu Asp Ala Asn Cys Ser Arg Asp Val Lys Gln Met Leu Lys 245 250 255  
 20 Leu Val Glu Leu Arg Ser Ser Asn Trp Gly Arg Val His Ala Thr Ser 260 265 270  
 Thr Tyr Arg Glu Ala Thr Pro Glu Asn Asp Pro Asn Tyr Phe Met Asn 275 280 285  
 25 Glu Pro Thr Phe Thr Thr Ser Asp Gly Val Pro Phe Thr Ala Ala Asp 290 295 300  
 Pro Asp Tyr Gln Glu Lys Tyr Gln Glu Leu Leu Glu Arg Glu Asp Phe 305 310 315  
 Phe Pro Asp Tyr Glu Glu Asn Gly Thr Asp Leu Ser Gly Ala Gly Asp 320 325 330 335  
 35 Pro Tyr Leu Asp Asp Ile Asp Asp Glu Met Asp Pro Glu Ile Glu Glu 340 345 350  
 Ala Tyr Glu Lys Phe Cys Leu Glu Ser Glu Arg Lys Arg Lys Gln 355 360 365  
 40  
 (2) INFORMATION FOR SEQ ID NO: 548:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 77 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 548:  
 50 Met Leu Arg Leu Asp Ile Ile Asn Ser Leu Val Thr Thr Val Phe Met 5 10 15  
 55 Leu Ile Val Ser Val Leu Ala Leu Ile Pro Glu Thr Thr Thr Thr 20 25 30  
 Val Gly Gly Gly Val Phe Ala Leu Val Thr Ala Val Cys Leu Ala 35 40 45  
 60 Asp Gly Ala Leu Ile Tyr Arg Lys Leu Leu Phe Asn Pro Ser Gly Pro

50 55 60  
 Tyr Gln Lys Lys Pro Val His Glu Lys Lys Glu Val Leu 65 70 75  
 10 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 47 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 549:  
 15 Met Leu Lys Gln Val Met Phe Val Phe Ser Gly Met Gly Pro Arg Ser 1 5 10 15  
 His Cys Trp Gly Leu Pro Leu Ala Cys Gly Thr Phe Val Gln Gly His 20 25 30  
 20 Gln Ala Asp Ser Ser His Leu Leu Pro Leu Lys His Gln Gly Ala 35 40 45  
 25  
 (2) INFORMATION FOR SEQ ID NO: 550:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 168 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (x1) SEQUENCE DESCRIPTION: SEQ ID NO: 550:  
 35 Met Leu Leu Ser Leu Ala Ala Phe Ser Val Ile Ser Val Val Ser Tyr 1 5 10 15  
 Leu Ile Leu Ala Leu Ser Val Thr Ile Ser Phe Arg Ile Tyr Lys 20 25 30  
 40 Ser Val Ile Gln Ala Val Gln Lys Ser Glu Glu Gly His Pro Phe Lys 35 40 45  
 45 Ala Tyr Leu Asp Val Asp Ile Thr Leu Ser Ser Glu Ala Phe His Asn 50 55 60  
 Tyr Met Asn Ala Ala Met Val His Ile Asn Arg Ala Leu Lys Leu Ile 65 70 75 80  
 50 Ile Arg Leu Phe Leu Val Glu Asp Leu Val Asp Ser Leu Lys Leu Ala 85 90 95  
 Val Phe Met Trp Leu Met Thr Tyr Val Gly Ala Val Phe Asn Gly Ile 100 105 110  
 55 Thr Leu Leu Ile Leu Ala Glu Leu Ile Phe Ser Val Pro Ile Val 115 120 125  
 Tyr Glu Lys Tyr Lys Thr Gln Ile Asp His Tyr Val Gly Ile Ala Arg 130 135 140

Asp Gln Thr Lys Ser Ile Val Glu Lys Ile Gln Ala Lys Leu Pro Gly  
145 150 155 160

5 Ile Ala Lys Lys Ala Glu Xaa  
165

10 (2) INFORMATION FOR SEQ ID NO: 551:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 124 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 551:

Ser Val Pro Phe His Leu Leu Val Val Leu Arg Ser Arg Ala Val Arg  
1 5 10 15

20 Ala Arg Arg Arg Arg Glu Pro Arg Ser Leu Pro Arg Pro Gly Asp Glu  
20 25 30

25 Glu Leu Gln Leu Leu Cys Gly Ala Arg Ser Asp Phe Leu Glu Arg  
35 40 45

Cys Glu Glu Asp Trp Val Cys Leu Trp His His Ala Asp His Ala Ala  
50 55 60

30 Phe Pro Gly Ser Phe Gln Cys His Gln Cys Gly Phe Leu Pro His Pro  
65 70 75 80

35 Gly Ser Ser Leu Cys His His Gln Leu Gln Asp Leu Gln Val Arg His  
85 90 95

Pro Ser Cys Thr Glu Val Arg Arg Arg Pro Ser Ile Gln Ser Leu Pro  
100 105 110

40 Gly Arg Arg His Tyr Ser Val Leu Arg Ser Phe Pro  
115 120

(2) INFORMATION FOR SEQ ID NO: 552:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 552:

Met Val His Leu Leu Val Leu Ser Gly Ala Trp Gly Met Gln Met Trp  
1 5 10 15

55 Val Thr Phe Val Ser Gly Phe Leu Phe Arg Ser Leu Pro Arg His  
20 25 30

Thr Phe Gly Leu Val Gln Ser Lys Leu Phe Pro Phe Tyr Phe His Ile  
35 40 45

Ser Met Gly Cys Ala Phe Ile Asn Leu Cys Ile Leu Ala Ser Gln His  
50 55 60

5 Ala Trp Ala Gln Leu Thr Phe Trp Glu Ala Ser Gln Leu Tyr Leu Leu  
65 70 75 80

Phe Leu Ser Leu Thr Leu Ala Thr Val Asn Ala Arg Trp Leu Glu Pro  
85 90 95

10 Arg Thr Thr Ala Ala Met Trp Ala Leu Gln Thr Val Glu Lys Glu Arg  
100 105 110

Gly Leu Gly Gly Glu Val Pro Gly Ser His Gln Gly Pro Asp Pro Tyr  
115 120 125

15 Arg Gln Leu Arg Glu Lys Asp Pro Lys Tyr Ser Ala Leu Arg Gln Asn  
130 135 140

20 Phe Phe Arg Tyr His Gly Leu Ser Ser Leu Cys Asn Leu Gly Cys Val  
145 150 155 160

Leu Ser Asn Gly Leu Cys Leu Ala Gly Leu Ala Leu Glu Ile Arg Ser  
165 170 175

25 Leu

30 (2) INFORMATION FOR SEQ ID NO: 553:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 72 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 553:

Met Ala Phe Ile Leu Leu Phe Tyr Cys Leu Met Thr Phe Leu Ser Leu  
1 5 10 15

40 Glu Gln Asn Ser Ala Thr Val Glu Pro Ser Ser His Glu Ile Leu His  
20 25 30

45 Leu Leu Gln Asn Cys Phe Glu Leu Leu Arg Thr Ser Thr Ser Gln Cys  
35 40 45

Thr Glu Gly Ile Pro Cys Ala Lys Ile Pro Glu Trp Val Thr His Leu  
50 55 60

50 Thr Trp Gln Thr Leu Lys Asn Ser  
65 70

55 (2) INFORMATION FOR SEQ ID NO: 554:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 45 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

60

680

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 554:

5 Val Leu Arg Ile Ile Cys Leu Trp Pro Cys Gly Thr Thr Leu Pro Leu 15  
1  
Val Glu Lys Ala His Asp Ser His Ser Ala Asp Pro Val Cys Pro Gly 30  
20  
Leu Thr Ala His Leu Pro Val Leu Leu Tyr Val Gln Leu 45  
35 40 45

(2) INFORMATION FOR SEQ ID NO: 555:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 251 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 555:

20 Met Lys His Ala Asp Pro Arg Ile Gln Gly Tyr Pro Leu Met Gly Ser 15  
1 5 10 15

25 Pro Leu Leu Met Thr Ser Ile Leu Leu Thr Tyr Val Tyr Phe Val Leu 30  
20 25 30

Ser Leu Gly Pro Arg Ile Met Ala Asn Arg Lys Pro Phe Gln Leu Arg 45  
35 40 45

30 Gly Phe Met Ile Val Tyr Asn Phe Ser Leu Val Ala Leu Ser Leu Tyr 60  
50 55 60

35 Ile Val Tyr Glu Phe Leu Met Ser Gly Trp Leu Ser Thr Tyr Thr Trp 80  
65 70 75 80

Arg Cys Asp Pro Gln Asp Cys Thr Leu Gly Gln Cys Pro Ser Val Pro 95  
85 90 95

40 Ser Pro Xaa Thr Pro Val Thr Lys Ala Tyr Val Val Arg Thr Glu Gln 110  
100 105 110

Gly Thr Gly Pro Pro Leu Pro Thr Ala Ala Leu Gln Gly Pro Arg Leu 125  
115 120 125

45 Trp Phe Leu Thr His Phe Pro Arg Ala Ala Pro Gly Met Trp Pro His 140  
130 135 140

50 Cys Cys Leu Pro Leu Gln Ser Trp Gly Leu Lys Gly Leu Tyr Ser Tyr 160  
145 150 155 160

Phe Pro Leu Pro Ala Leu Lys Leu Gly Arg Gly Ala Leu Arg Ala Gly 175  
165 170 175

55 Pro Thr Lys Gly Leu Val Ala Phe Phe Leu Thr Gln Lys Arg Ser Ala 190  
180 185 190

Ile Met Ser Leu Trp Thr Gln Ser His Ser Thr Thr Pro His Thr Glu 205  
195 200 205

60

681

Ala Val Ala Ser Gly Pro Lys Val Arg Val Gly Gly Gly Leu Gly Ile 220  
210 215 220

5 Gln Pro Val Glu Ala Ala Tyr Ser Thr Cys Val Leu Ile Lys Ser Asp 240  
225 230 235 240

Arg Gly Asn His Glu Lys Lys Lys Lys Lys Lys 250  
245 250

(2) INFORMATION FOR SEQ ID NO: 556:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 19 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 556:

20 Gly Leu Ala Gly Leu Cys Gly Gln Leu Ser Ser Pro Ala Leu Cys Val 15  
1 5 10 15

Asn Arg Leu

25

(2) INFORMATION FOR SEQ ID NO: 557:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 217 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 557:

35 Met Ile Thr Glu Lys Trp Gly Leu Asn Met Glu Tyr Cys Arg Gly Gln 15  
1 5 10 15

Ala Tyr Ile Xaa Ser Ser Gly Phe Ser Ser Lys Met Lys Val Val Ala 30  
20 25 30

Ser Arg Leu Leu Glu Lys Tyr Pro Gln Ala Ile Tyr Thr Leu Cys Ser 45  
35 40 45

45 Ser Cys Ala Leu Asn Met Trp Leu Ala Lys Ser Val Pro Val Met Gly 60  
50 55 60

Val Ser Val Ala Leu Gly Thr Ile Glu Glu Val Cys Ser Phe Phe His 80  
65 70 75 80

50 Arg Ser Pro Gln Leu Leu Leu Glu Leu Asp Asn Val Ile Ser Val Leu 95  
85 90 95

Phe Gln Asn Ser Lys Glu Arg Gly Lys Glu Leu Lys Glu Ile Cys His 110  
100 105 110

Ser Gln Trp Thr Gly Arg His Asp Ala Phe Glu Ile Leu Val Glu Leu 125  
115 120 125

60 Leu Gln Ala Leu Val Leu Cys Leu Asp Gly Ile Asn Ser Asp Thr Asn

130 135 140  
 Ile Arg Trp Asn Asn Tyr Ile Ala Gly Arg Ala Phe Val Leu Cys Ser  
 145 150 155 160  
 5 Ala Val Ser Asp Phe Asp Phe Ile Val Thr Ile Val Val Leu Lys Asn  
 165 170 175  
 10 Val Leu Ser Phe Thr Arg Ala Phe Gly Lys Asn Leu Gln Gly Gln Thr  
 180 185 190  
 Ser Asp Val Phe Phe Ala Ala Gly Ser Leu Thr Ala Val Leu His Ser  
 195 200 205  
 15 Leu Asn Glu Val Ile Gly Lys Tyr Xaa  
 210 215  
 20 (2) INFORMATION FOR SEQ ID NO: 558:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 82 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 558:  
 Leu Leu Lys Val Leu Cys Ile Leu Pro Val Met Lys Val Glu Asn Glu  
 1 5 10 15  
 30 Arg Tyr Glu Asn Gly Arg Lys Arg Leu Lys Ala Tyr Leu Arg Asn Thr  
 20 25 30  
 35 Leu Thr Asp Gln Arg Ser Ser Asn Leu Ala Leu Leu Asn Ile Asn Phe  
 35 40 45  
 Asp Ile Lys His Asp Leu Asp Leu Met Val Asp Thr Tyr Ile Lys Leu  
 50 55 60  
 40 Tyr Thr Ser Lys Ser Glu Leu Pro Thr Asp Asn Ser Glu Thr Val Glu  
 65 70 75 80  
 Asn Thr  
 45  
 (2) INFORMATION FOR SEQ ID NO: 559:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 95 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 559:  
 Met Val Leu Ile Leu Leu Asn Leu Leu Leu Gly Gln Phe Ser Cys Met  
 1 5 10 15  
 55 Ser Pro Ala Ser His His Cys His Pro Leu Pro Thr Glu Met Pro Cys  
 20 25 30 60

Ser Ser Asp Trp Gly Phe Asp Ser His Thr Val Tyr Pro Ser Cys Val  
 35 40 45  
 5 Asp Ala Leu Leu Pro Lys Pro Ser Ala Asn Ser Phe Pro Asn Gly Ser  
 50 55 60  
 Cys His Cys Gln Gly Leu Tyr Asn Gln Gln Gln Asn Leu His Ala  
 65 70 75 80  
 10 Ala Glu Gly Pro Ala Ser Leu Arg Cys Asn Lys Tyr Val Ser Thr  
 85 90 95  
 15 (2) INFORMATION FOR SEQ ID NO: 560:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 54 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 560:  
 Met Ile Pro Ala Tyr Ser Lys Asn Arg Ala Tyr Ala Ile Phe Phe Ile  
 1 5 10 15  
 Val Phe Thr Val Ile Gly Asp Ala Pro Gly Ala Val Leu Ser Cys Ala  
 20 25 30  
 30 Gly His Pro Cys Val Gly Phe Ala Ala Val Leu Val Ala Pro Leu Thr  
 35 40 45  
 Val Ala Val Ser Ser Xaa  
 50  
 35  
 (2) INFORMATION FOR SEQ ID NO: 561:  
 (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 108 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 561:  
 Met Glu Val Pro Pro Ala Pro Arg Ser Phe Leu Cys Arg Ala Leu  
 1 5 10 15  
 Cys Leu Phe Pro Arg Val Phe Ala Ala Glu Ala Val Thr Ala Asp Ser  
 20 25 30  
 Glu Val Leu Glu Glu Arg Gln Lys Arg Leu Pro Tyr Val Pro Glu Pro  
 35 40 45  
 55 Tyr Tyr Pro Glu Ser Gly Tyr Asp Arg Leu Arg Glu Leu Phe Gly Lys  
 50 55 60  
 Asp Thr Val Asn Thr Ser Leu Asn Val Tyr Arg Asn Lys Asp Ala Leu  
 65 70 75 80

Ser His Phe Val Ile Ala Gly Ala Val Thr Gly Ser Leu Phe Arg Ile  
85 90 95

5 Asn Val Gly Leu Arg Gly Trp Trp Leu Val Ala Xaa  
100 105

10 (2) INFORMATION FOR SEQ ID NO: 562:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 50 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 562:

15 Met Asn Trp Gly Leu Ser Ile Trp Leu His Tyr Gly Lys Lys Lys  
1 5 10 15

20 Glu Gln Val Phe Leu Val Ile Leu Ala His Val Val Arg Arg Cys Ala  
20 25 30

Ser Asp Gly Ile Leu Gln Phe Gly Ser Ser Leu Leu Lys Met Arg Arg  
35 40 45

25 Ala Pro  
50

30 (2) INFORMATION FOR SEQ ID NO: 563:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 253 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 563:

40 Met Val Lys Val Cys Asn Asp Ser Asp Arg Trp Ser Leu Ile Ser Leu  
1 5 10 15

Ser Asn Asn Ser Gly Lys Asn Val Glu Leu Lys Phe Val Asp Ser Leu  
20 25 30

45 Arg Arg Gln Phe Glu Phe Ser Val Asp Ser Phe Gln Ile Lys Leu Asp  
35 40 45

Ser Leu Leu Leu Phe Tyr Glu Cys Ser Glu Asn Pro Met Thr Glu Thr  
50 55 60

50 Phe His Pro Thr Ile Ile Gly Glu Ser Val Tyr Gly Asp Phe Gln Glu  
65 70 75

55 Ala Phe Asp His Leu Cys Asn Lys Ile Ile Ala Thr Arg Asn Pro Glu  
85 90 95

Glu Ile Arg Gly Gly Glu Leu Lys Tyr Cys Asn Leu Val Arg  
100 105 110

60 Gly Phe Arg Pro Ala Ser Asp Glu Ile Lys Thr Leu Gln Arg Tyr Met

115 120 125

5 Cys Ser Arg Phe Phe Ile Asp Phe Ser Asp Ile Gly Glu Gln Arg  
130 135 140

Lys Leu Glu Ser Tyr Leu Gln Asn His Phe Val Gly Leu Glu Asp Arg  
145 150 155 160

10 Lys Tyr Glu Tyr Leu Met Thr Leu His Gly Val Val Asn Glu Ser Thr  
165 170 175

Val Cys Leu Met Gly His Glu Arg Arg Gln Thr Leu Asn Leu Ile Thr  
180 185 190

15 Met Leu Ala Ile Arg Val Leu Ala Asp Gln Asn Val Ile Pro Asn Val  
195 200 205

20 Ala Asn Val Thr Cys Tyr Tyr Gln Pro Ala Pro Tyr Val Ala Asp Ala  
210 215 220

Asn Phe Ser Asn Tyr Tyr Ile Ala Gln Val Gln Pro Val Phe Thr Cys  
225 230 235 240

25 Gln Gln Gln Thr Tyr Ser Thr Trp Leu Pro Cys Asn Xaa  
245 250

30 (2) INFORMATION FOR SEQ ID NO: 564:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 564:

Met Ser Phe Leu Met Trp Leu Met Ser Leu Ala Ile Thr Ser Gln Pro  
1 5 10 15

40 Pro Met

45 (2) INFORMATION FOR SEQ ID NO: 565:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 80 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 565:

Met Ala Pro Lys Gly Lys Val Gly Thr Arg Gly Lys Lys Gln Ile Phe  
1 5 10 15

50 Glu Glu Asn Arg Glu Thr Leu Lys Phe Tyr Leu Arg Ile Ile Leu Gly  
20 25 30

60 Ala Asn Ala Ile Tyr Cys Leu Val Thr Leu Val Phe Phe Tyr Ser Ser  
35 40 45

686

Ala Ser Phe Trp Ala Trp Leu Ala Leu Gly Phe Ser Leu Ala Val Tyr  
50 55 60

5 Gly Ala Ser Tyr His Ser Met Ser Ser Met Ala Arg Ala Ala Phe Phe  
65 70 75 80

10

(2) INFORMATION FOR SEQ ID NO: 566:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 73 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 566:

20 His Leu Lys Asp Val Ile Leu Leu Thr Ala Ile Val Gln Val Leu Ser  
1 5 10 15

25 Cys Phe Ser Leu Tyr Val Trp Ser Phe Trp Leu Leu Ala Pro Gly Arg  
20 25 30

Ala Leu Tyr Leu Leu Trp Val Asn Val Leu Gly Pro Trp Phe Thr Ala  
35 40 45

30 Asp Ser Gly Thr Pro Ala Pro Glu His Asn Glu Lys Arg Gln Arg Arg  
50 55 60

35 Gln Glu Arg Arg Gln Met Lys Arg Leu  
65 70

(2) INFORMATION FOR SEQ ID NO: 567:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 263 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 567:

45 Met Asp Cys Pro Ala Leu Pro Pro Gly Trp Lys Lys Glu Glu Val Ile  
1 5 10 15

50 Arg Lys Ser Gly Leu Ser Ala Gly Lys Ser Asp Val Tyr Tyr Phe Ser  
20 25 30

Pro Ser Gly Lys Lys Phe Arg Ser Lys Pro Gln Leu Ala Arg Tyr Leu  
35 40 45

55 Gly Asn Thr Val Asp Leu Ser Ser Phe Asp Phe Arg Thr Gly Lys Met  
50 55 60

Met Pro Ser Lys Leu Gln Lys Asn Lys Gln Arg Leu Arg Asn Asp Pro  
65 70 75 80

687

Leu Asn Gln Asn Lys Gly Lys Pro Asp Leu Asn Thr Thr Leu Pro Ile  
85 90 95

Arg Gln Thr Ala Ser Ile Phe Lys Gln Pro Val Thr Lys Val Thr Asn  
100 105 110

His Pro Ser Asn Lys Val Lys Ser Asp Pro Gln Arg Met Asn Glu Gln  
115 120 125

10 Pro Arg Gln Leu Phe Trp Glu Lys Arg Leu Gln Gly Leu Ser Ala Ser  
130 135 140

15 Asp Val Thr Glu Gln Ile Ile Lys Thr Met Glu Leu Pro Lys Gly Leu  
145 150 155 160

Gln Gly Val Gly Pro Gly Ser Asn Asp Glu Thr Leu Leu Ser Ala Val  
165 170 175

20 Ala Ser Ala Leu His Thr Ser Ser Ala Pro Ile Thr Gly Gln Val Ser  
180 185 190

Ala Ala Val Glu Lys Asn Pro Ala Val Trp Leu Asn Thr Ser Gln Pro  
195 200 205

25 Leu Cys Lys Ala Phe Ile Val Thr Asp Glu Asp Ile Arg Lys Gln Glu  
210 215 220

Glu Arg Val Gln Gln Val Arg Lys Lys Leu Glu Glu Ala Leu Met Ala  
225 230 235 240

30 Asp Ile Leu Ser Arg Ala Ala Asp Thr Glu Glu Met Asp Ile Glu Met  
245 250 255

35 Asp Ser Gly Asp Glu Ala Xaa  
260

(2) INFORMATION FOR SEQ ID NO: 568:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 70 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 568:

Met Met Arg Pro Phe Tyr Leu Leu Leu Pro Val Leu Cys Thr Gln Ala  
1 5 10 15

50 Leu Arg Gln Ser Gln Gly Lys Ser Pro Leu Leu Trp Lys Arg Thr Leu  
20 25 30

Leu Phe Gly Leu Thr His Leu Asn Pro Ser Ala Lys Leu Leu Ser  
35 40 45

55 Gln Met Lys Thr Ser Gly Asn Arg Lys Ser Glu Tyr Ser Lys Tyr Ala  
50 55 60

Arg Asn Trp Lys Lys His  
65 70



688

WO 98/39448

689

(2) INFORMATION FOR SEQ ID NO: 569:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 34 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 569:

Met Pro Val Thr Ser Lys Arg Thr Leu Phe Phe Pro Asp Pro Cys Ser  
1 5 10 15

Tyr Asp Thr Pro Pro Asp Cys His Cys His Ser Phe Arg Ala Glu  
20 25 30

Leu Leu

(2) INFORMATION FOR SEQ ID NO: 570:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 104 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 570:

Met Asn Ser Arg Gly Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu  
1 5 10 15

Leu His Ile Val Leu Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala  
20 25 30

Trp Thr Leu Thr Asn Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu  
35 40 45

His Ala Val Lys Gly Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ser  
50 55 60

Lys Ala Pro Asn Ser Leu Gly Thr Thr Gly Leu Trp Ser Thr Val Tyr  
65 70 75 80

Ile Phe Thr Glu Val Phe His Asn Phe Ser Asn Asn Ser Ile Phe Ser  
85 90 95

Gly Lys Phe Leu Tyr Glu Val Asn  
100

(2) INFORMATION FOR SEQ ID NO: 571:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 132 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 571:

Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser  
1 5 10 15

Met Trp Leu Thr Tyr Ala Leu Gly Val Gly Leu Leu His Ile Val Leu  
1 5 10 15

Leu Ser Ile Pro Phe Phe Ser Val Pro Val Ala Trp Thr Leu Thr Asn  
20 25 30

Ile Ile His Asn Leu Gly Met Tyr Val Phe Leu His Ala Val Lys Gly  
35 40 45

Thr Pro Phe Glu Thr Pro Asp Gln Gly Lys Ala Arg Leu Leu Thr His  
50 55 60

Trp Glu Gln Leu Asp Tyr Gly Val Gln Phe Thr Ser Ser Arg Lys Phe  
65 70 75 80

Phe Thr Ile Ser Pro Ile Ile Leu Tyr Phe Leu Ala Ser Phe Tyr Thr  
85 90 95

Lys Tyr Asp Pro Thr His Phe Ile Leu Asn Thr Ala Ser Leu Leu Ser  
100 105 110

Val Leu Ile Pro Lys Met Pro Gln Leu His Gly Val Arg Ile Phe Gly  
115 120 125

Ile Asn Lys Tyr  
130

(2) INFORMATION FOR SEQ ID NO: 572:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 572:

Met Asn Lys Trp Ile Cys Glu Met His Cys Tyr Leu Val Leu Leu Ser  
1 5 10 15

Val Cys Ser Pro Ser Ala Leu Arg Arg Val Arg His Thr Leu Ser Arg  
20 25 30

(2) INFORMATION FOR SEQ ID NO: 573:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 28 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 573:

Met Pro Val Leu Ser Leu Leu Cys Thr Leu Ile Val Ser Phe Gln Ser  
1 5 10 15

Ala Asp Ser Cys Glu Val Phe Leu Asn Cys Ser Leu  
20 25

5

(2) INFORMATION FOR SEQ ID NO: 574:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 40 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 574:

Met Lys Val Ser Thr Met Leu Trp Phe Leu Cys Trp Glu Gln Ser His  
1 5 10 15

Phe Leu Arg Glu Trp Glu Asp Leu Ser Thr Phe Leu Ile Leu Ile Gln  
20 25 30

Met Glu Cys Gln Tyr Gly Asn Ser  
35 40

25

(2) INFORMATION FOR SEQ ID NO: 575:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 575:

Met Gly Leu Pro Leu Met Ala Leu Met Trp Ser Thr Leu Pro Ala Ser  
1 5 10 15

Ala Gly Val Asn Phe Ile Leu Ala Leu Pro Leu Leu Xaa Leu  
20 25 30

40

(2) INFORMATION FOR SEQ ID NO: 576:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 576:

Met Lys Arg Gly Cys Leu Gly Leu Phe Phe Ser Cys Cys Ser Ser  
1 5 10 15

Ala Pro Thr Met Leu Leu Cys Asp Tyr Leu Asn Trp Phe  
20 25

55

(2) INFORMATION FOR SEQ ID NO: 577:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

Thr Phe Thr Ser Asp Phe Tyr Phe Met Glu Phe Gly Ile Glu Val Lys  
35 40 45

60

(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 577:

Met Lys Leu Leu Leu Gly Ile Ala Leu Leu Ala Tyr Val Ala Ser Val  
1 5 10 15

Trp Gly Asn Phe Val Asn Met Arg Ser Ile Gln Glu Asn Gly Glu Leu  
20 25 30

Lys Ile Glu Ser Lys Ile Glu Glu Met Val Glu Pro Leu Arg Glu Lys  
35 40 45

Ile Arg Asp Leu Glu Lys Ser Phe Thr Gln Lys Tyr Pro Pro Val Lys  
50 55 60

Phe Leu Ser Glu Lys Asp Arg Lys Arg Ile Leu Xaa Asn Arg Arg Arg  
65 70 75 80

Xaa Val Arg Gly Leu Pro Ser Xaa Leu Thr Asn Ser  
85 90

25

(2) INFORMATION FOR SEQ ID NO: 578:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 42 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 578:

Met Lys Phe Ser Leu Val Leu Leu Ile Lys Ile Ile Ser Phe Glu Arg  
1 5 10 15

Leu Leu Ile Phe Leu Phe Pro Leu Ser Phe Leu Pro Asn Ile Trp Arg  
20 25 30

Arg Val Met Val Asn Leu Asn Ile Leu Phe  
35 40

45

(2) INFORMATION FOR SEQ ID NO: 579:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 70 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 579:

Leu Ala Gln Glu Cys Pro Pro His Ile Pro Ser Ser Phe Phe Leu Val  
1 5 10 15

Lys Leu Leu Phe Ile Pro Trp Leu Ala Ser Leu Leu Pro Leu Ser  
20 25 30

Thr Phe Thr Ser Asp Phe Tyr Phe Met Glu Phe Gly Ile Glu Val Lys  
35 40 45

60

Leu Gln Gln Cys Arg Gln His Gln Val Leu Gln Glu Lys Asn Thr Lys  
50 55 60

Lys Phe Asn Lys Lys Lys  
65 70

10 (2) INFORMATION FOR SEQ ID NO: 580:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 110 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

15 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 580:

Met Leu Arg Leu Leu Leu Val Ala Phe Ala Leu Val Val Val Leu  
1 5 10 15

20 Phe His Val Leu Leu Ala Pro Ile Thr Ala Leu Phe His Thr His Phe  
20 25 30

Tyr Asp Arg Leu Leu Asp Ala Gly Ser Arg Trp Pro Glu Leu Tyr Leu  
35 40 45

25 Tyr Ser Arg Ala Asp Glu Val Val Leu Ala Arg Asp Ile Glu Arg Met  
50 55 60

30 Val Glu Ala Arg Leu Ala Arg Arg Val Leu Ala Arg Ser Val Asp Phe  
65 70 75 80

Val Ser Ser Ala His Val Ser His Leu Arg Asp Tyr Pro Thr Tyr Tyr  
85 90 95

35 Thr Ser Leu Cys Val Asp Phe Met Arg Asn Cys Val Arg Cys  
100 105 110

40 (2) INFORMATION FOR SEQ ID NO: 581:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 30 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 581:

Met Phe Lys Leu Glu Glu Cys Gly Lys Thr Thr Phe Leu Leu Ser Met  
1 5 10 15

50 Ala Leu Tyr Phe Trp Trp Ile Val Gln Thr Thr Lys Gly Cys  
20 25 30

55 (2) INFORMATION FOR SEQ ID NO: 582:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 71 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 582:

5 Met Glu Ser Asp Ala Leu Leu Thr Ile Phe Trp Ile Ile Ala Arg  
1 5 10 15

Ser Ser Val Arg Ser Val Gly Lys Ser Ser Gln Arg Ser Phe Thr Thr  
20 25 30

10 Ile Thr Gln Leu Arg Ser Thr His Thr Gly Pro Ser Arg Arg Ser Tyr  
35 40 45

Leu Ile Trp Trp Asn Gly Gly Pro Lys Arg Thr Ile Ser Tyr Val Ser  
50 55 60

15 Arg Arg Phe Arg Ser Phe Arg  
65 70

20 (2) INFORMATION FOR SEQ ID NO: 583:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 47 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 583:

30 Val Gly Leu Phe Gln Pro Lys Thr Phe Gln Val Pro Val Thr Asp Leu  
1 5 10 15

Tyr Ile Phe Ile Lys Ile Tyr Ser Glu Ile Gly Pro Ile Met His Val  
20 25 30

35 Leu Cys Pro Gly Tyr Ser Gln Ser Pro Ser Thr Pro Pro Trp Thr  
35 40 45

40 (2) INFORMATION FOR SEQ ID NO: 584:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 39 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 584:

Met Trp Phe Gly Ser Asp Arg Ser Asp Leu Arg Ile Gly Thr Ala Phe  
1 5 10 15

50 Leu Phe Asp Leu Val Cys Asp Leu Cys Ile His Ala Trp Lys Pro Pro  
20 25 30

55 Gly Leu Val Arg Phe Ser Phe  
35

60 (2) INFORMATION FOR SEQ ID NO: 585:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 585:

Met Leu Asn Thr Ala Ser Leu Asn Leu Pro Trp Lys Val Gln Leu Phe  
1 5 10 15

10 Ala His Ala

15 (2) INFORMATION FOR SEQ ID NO: 586:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 23 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 586:

Met Ser Ala Cys Leu Leu Phe Leu Ala Phe Ser Trp Lys Arg Lys  
1 5 10 15

25 Gly Leu Trp Ser Gly Pro Gly  
20

30 (2) INFORMATION FOR SEQ ID NO: 587:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 69 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 587:

Met Leu Pro Pro Phe Ser Leu Val Tyr Thr His Phe Leu Val Ala Ser  
1 5 10 15

Leu Leu Pro Val Ile Leu Ala Val Phe Pro Asp Ser Ala Gln Ile Val  
20 25 30

Pro Leu Leu Lys Pro Ile Pro Arg Pro Gln Pro Glu Val Ile Phe Pro  
35 40 45

Ser Ser Glu Leu Leu Gln Leu Leu Ser Val Gln Phe Val Trp Gln  
50 55 60

50 Ala His Thr Val Ala  
65

55 (2) INFORMATION FOR SEQ ID NO: 588:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 77 amino acids

(B) TYPE: amino acid

## (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 588:

Met Gly Pro Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu  
5 10 15

Leu Met Met Gly Ala Ile Phe Thr Leu Ala Ala Leu Lys Glu Ser Leu  
20 25 30

Ser Thr Cys Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Leu  
35 40 45

Asn Val Gly Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr  
50 55 60

Arg Lys Lys Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys  
65 70 75

20 (2) INFORMATION FOR SEQ ID NO: 589:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 155 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 589:

Met Ala Leu Leu Leu Ser Val Leu Arg Val Leu Leu Gly Gly Phe Phe  
1 5 10 15

Ala Leu Val Gly Leu Ala Lys Leu Ser Glu Glu Ile Ser Ala Pro Val  
20 25 30

Ser Glu Arg Met Asn Ala Leu Phe Val Gln Phe Ala Glu Val Phe Pro  
35 40 45

Leu Lys Val Phe Gly Tyr Gln Pro Asp Pro Leu Asn Tyr Gln Ile Ala  
50 55 60

Val Gly Phe Leu Glu Leu Leu Ala Gly Leu Leu Val Met Gly Pro  
65 70 75 80

Pro Met Leu Gln Glu Ile Ser Asn Leu Phe Leu Ile Leu Leu Met Met  
85 90 95

Gly Ala Ile Phe Thr Leu Ala Leu Lys Glu Ser Leu Ser Thr Cys  
100 105 110

Ile Pro Ala Ile Val Cys Leu Gly Phe Leu Leu Leu Asn Val Gly  
115 120 125

Gln Leu Leu Ala Gln Thr Lys Lys Val Val Arg Pro Thr Arg Lys Lys  
130 135 140

Thr Leu Ser Thr Phe Lys Glu Ser Trp Lys Xaa  
145 150 155

60

## (2) INFORMATION FOR SEQ ID NO: 590:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 24 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 590:

Met Pro Glu Thr Arg Leu Gly His Arg Gln Gln Phe Ala Val Phe His  
 1 5 10 15

Leu Xaa Pro Val Pro Pro Cys Gly  
 20

15

## (2) INFORMATION FOR SEQ ID NO: 591:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 38 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 591:

Met Leu Thr Phe Leu Phe Ser Ala Cys Ala Thr Cys Leu Gly Lys Leu  
 1 5 10 15

Ala Ser Pro Leu Ala Pro Val Gly Pro Gln Gln Arg Gly Xaa Pro Pro  
 20 25 30

Gly Pro Pro Leu Leu Ser  
 35

30

35

## (2) INFORMATION FOR SEQ ID NO: 592:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 69 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 592:

Met Asp Pro Phe His Tyr Asp Tyr Gln Thr Leu Arg Ile Gly Gly Leu  
 1 5 10 15

Val Phe Ala Val Val Leu Phe Ser Val Gly Ile Leu Leu Ile Leu Ser  
 20 25 30

Arg Arg Cys Lys Cys Ser Phe Asn Gln Lys Pro Arg Ala Pro Gly Asp  
 35 40 45

Glu Glu Ala Gln Val Glu Asn Leu Ile Thr Ala Asn Ala Thr Glu Pro  
 50 55 60

Gln Lys Ala Glu Asn  
 65

55

60

## (2) INFORMATION FOR SEQ ID NO: 593:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 308 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 593:

Asn Leu Arg Val Arg Leu Gly Asp Val Ile Ser Ile Gln Pro Cys Pro  
 1 5 10 15

Asp Val Lys Tyr Gly Lys Arg Ile His Val Leu Pro Ile Asp Asp Thr  
 20 25 30

Val Glu Gly Ile Thr Gly Asn Leu Leu Phe Glu Val Tyr Leu Lys Pro Tyr  
 35 40 45

Phe Leu Glu Ala Tyr Arg Pro Ile Arg Lys Gly Asp Ile Phe Leu Val  
 50 55 60

Arg Gly Gly Met Arg Ala Val Glu Phe Lys Val Val Glu Thr Asp Pro  
 65 70 75 80

Ser Pro Tyr Cys Ile Val Ala Pro Asp Thr Val Ile His Cys Glu Gly  
 85 90 95

Glu Pro Ile Lys Arg Glu Asp Glu Glu Ser Leu Asn Glu Val Gly  
 100 105 110

Tyr Asp Asp Ile Gly Gly Cys Arg Lys Gln Leu Ala Gln Ile Lys Glu  
 115 120 125

Met Val Glu Leu Pro Leu Arg His Pro Ala Leu Phe Lys Ala Ile Gly  
 130 135 140

Val Lys Pro Pro Arg Gly Ile Leu Leu Tyr Gly Pro Pro Gly Thr Gly  
 145 150 155 160

Lys Thr Leu Ile Ala Arg Ala Val Ala Asn Glu Thr Gly Ala Phe Phe  
 165 170 175

Phe Leu Ile Asn Gly Pro Glu Ile Met Ser Lys Leu Ala Gly Glu Ser  
 180 185 190

Glu Ser Asn Leu Arg Lys Ala Phe Glu Glu Ala Glu Lys Asn Ala Pro  
 195 200 205

Ala Ile Ile Phe Ile Asp Glu Leu Asp Ala Ile Ala Pro Lys Arg Glu  
 210 215 220

Lys Thr His Gly Glu Val Glu Arg Arg Ile Val Ser Gln Leu Leu Thr  
 225 230 235 240

Leu Met Asp Gly Leu Lys Gln Arg Ala His Val Ile Val Met Ala Ala  
 245 250 255

Thr Asn Arg Pro Asn Ser Ile Asp Pro Ala Leu Arg Arg Phe Gly Arg  
 260 265 270

Phe Asp Arg Glu Val Asp Ile Gly Ile Pro Asp Ala Thr Gly Arg Leu  
 280 285 290

698

275 280 285

Glu Ile Leu Gln Ile His Thr Lys Asn Met Lys Leu Ala Asp Asp Val  
290 295 300

5

Asp Leu Glu Gln  
305

10

(2) INFORMATION FOR SEQ ID NO: 594:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 22 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 594:

Met Gln Ile Lys Leu Lys Ser Val Lys Thr Val Phe Ala Ile Thr  
1 5 10 15

Leu Val Leu Phe Leu  
20

25

(2) INFORMATION FOR SEQ ID NO: 595:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 595:

Met Phe Pro Lys Phe Cys Pro Ile Leu Ser Leu Val Asp Phe Ile Ser  
1 5 10 15

His Arg Asp Lys Pro Glu Thr Glu  
20

40

(2) INFORMATION FOR SEQ ID NO: 596:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 596:

Met Leu Ile Glu Cys Ala Trp Gln Leu Met Phe Leu Leu Lys Val  
1 5 10 15

Glu Gln Leu Gly Ile Leu Asp Lys  
20

55

(2) INFORMATION FOR SEQ ID NO: 597:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids

60

699

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 1 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 597:

5

Met  
1

10

(2) INFORMATION FOR SEQ ID NO: 598:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 8 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 598:

Met Cys Ile Met Ser Ala Leu Val  
1 5

25

(2) INFORMATION FOR SEQ ID NO: 599:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 25 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 599:

Met Phe Leu Val Trp Phe Trp Gly Leu Ile Ser Ala Leu Ser Asn  
1 5 10 15

Val His Thr Pro Ser Arg Leu Pro Ala  
20 25

35

40

(2) INFORMATION FOR SEQ ID NO: 600:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 27 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 600:

Met Xaa Gly Leu Ser Leu Ile Leu Thr Val Thr Leu Leu Ala Val Ser  
1 5 10 15

Asp Ser Ala Ala Thr Cys Ile Val Ala Lys Gly  
20 25

55

(2) INFORMATION FOR SEQ ID NO: 601:

- (i) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 61 amino acids

60

700

(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 601:

5 Met Trp Thr Arg Ser Ser Arg Cys Leu Leu Cys Ile Pro Gly Xaa  
1 5 10 15

Ser Arg Arg Arg Ala Gly Ser Gly Met Lys Pro Arg Ser Trp Ser  
20 25 30

10 Ala Trp Arg Pro Ser Gly Gly Thr Gly Thr Ser Ser Gln Ser Ser  
35 40 45

Thr Gln Ser Arg Thr Leu Ser Ala Thr Ala Ser Pro Ala  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 602:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 602:

25 Met Arg Glu Thr Ser Ile Arg Val Leu Leu Met Leu Pro Ala Leu Glu  
1 5 10 15

30 Ser Thr Ser Gly Leu Ser Ala Phe Met Gly Leu Gly Thr  
20 25

(2) INFORMATION FOR SEQ ID NO: 603:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 69 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 603:

45 Met Pro Pro Lys Gln Glu Leu Gly Ser Gly Val Gly Glu Leu Ala Lys  
1 5 10 15

Ann Ser Lys Arg Gln His Trp Ann His Arg Trp Lys Lys Tyr Leu Lys  
20 25 30

50 Leu Ile Arg Trp Glu Asp Gly Leu Leu Glu Gly Leu Leu Val  
35 40 45

Leu Glu His Cys Ala Thr Met Ala Trp Asp Cys Leu Met Arg Leu Glu  
50 55 60

55 Leu Leu Lys Arg Leu  
65

(2) INFORMATION FOR SEQ ID NO: 604:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 24 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 604:

5 Lys Ile Val Tyr Ile Leu Gly Ann Pro Leu Lys Phe Ann Ser Arg Val  
1 5 10 15

10 Ile His His Leu Val Leu Leu Gln  
20

(2) INFORMATION FOR SEQ ID NO: 605:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 35 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 605:

15 Met Ann Leu His Gln Arg Arg Leu Leu Ile Gly His Leu Met Thr  
1 5 10 15

25 Leu Val Lys Ala Ser Lys Ser Phe Thr Glu Ile Thr Ser Ser  
20 25 30

30 Arg Lys Lys  
35

(2) INFORMATION FOR SEQ ID NO: 606:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 130 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 606:

45 Leu Leu Gly Tyr Gly Leu Phe Gly His Cys Ile Val Leu Phe Ile Thr  
1 5 10 15

Tyr Ann Ile His Leu His Ala Leu Phe Tyr Leu Phe Trp Leu Leu Val  
20 25 30

50 Gly Gly Leu Ser Thr Leu Arg Met Val Ala Val Leu Val Ser Arg Thr  
35 40 45

Val Gly Pro Thr Gln Arg Leu Leu Cys Gly Thr Leu Ala Ala Leu  
50 55 60

55 His Met Leu Phe Leu Leu Tyr Leu His Phe Ala Tyr His Lys Val Xaa  
65 70 75 80

Glu Gly Ile Leu Asp Thr Leu Glu Gly Pro Ann Ile Pro Pro Ile Gln  
85 90 95

60

702

Arg Val Pro Arg Asp Ile Pro Ala Met Leu Pro Ala Ala Arg Leu Pro  
100 105 110

Thr Thr Val Leu Asn Ala Thr Ala Lys Ala Val Ala Val Thr Leu Gln  
115 120 125

Ser His  
130

10

(2) INFORMATION FOR SEQ ID NO: 607:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 23 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 607:

Met Leu Val Ile Phe Leu Phe Thr Ser Leu Leu Lys Ile Pro Ser Ser  
1 5 10 15

Val Pro Gly Leu Ile Asn Val  
20

25

(2) INFORMATION FOR SEQ ID NO: 608:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 6 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 608:

Glu Leu Asp Tyr Ile Leu  
1 5

35

(2) INFORMATION FOR SEQ ID NO: 609:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 232 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 609:

Met Ala Pro Pro Gly Tyr Gln Xaa Xaa Xaa Xaa Trp Leu Ala Cys  
1 5 10 15

Pro Asp Arg Gly Glu Leu Ser Ser Arg Ser Pro Pro Cys Arg Leu Ala  
20 25 30

55

Arg Trp Ala Glu Gly Asp Arg Glu Thr Arg Thr Cys Leu Leu Glu Leu  
35 40 45

Ser Ala Gln Ser Trp Gly Gly Arg Phe Arg Arg Ser Ser Ala Val Ser  
50 55 60

60

703

Ala Gly Ser Pro Ser Arg Leu His Phe Leu Pro Gln Pro Leu Leu Leu  
65 70 75 80

Arg Ser Ser Gly Ile Pro Ala Ala Thr Pro Trp Pro Gln Pro Ala  
85 90 95

Gly Leu Pro Val Arg Pro Thr Arg Thr Gly Glu Asp Arg  
100 105 110

Thr Leu Asp Ile Ser Ile Cys Thr Glu Val Leu Ala Gly Thr Glu Gln  
115 120 125

Pro Pro Pro Arg Met Thr Ser Pro Ser Ser Ser Pro Val Phe Arg  
130 135 140

Leu Glu Thr Leu Asp Gly Gly Gln Glu Asp Gly Ser Glu Ala Asp Arg  
145 150 155 160

Gly Lys Leu Asp Phe Gly Ser Gly Leu Pro Pro Met Glu Ser Gln Phe  
165 170 175

Gln Gly Glu Asp Arg Lys Phe Ala Pro Ser Asp Lys Ser Gln Pro Pro  
180 185 190

Thr Thr Glu Arg Glu Gln Val Pro Val Ser Arg Ile Gln Thr Asp Leu  
195 200 205

Thr Glu Ile Gly Ser Ser Met Arg Ser Pro Gly Val Ser Pro Arg Ile  
210 215 220

Trp Leu Asp Phe Gln Ser Thr Xaa  
225 230

35

(2) INFORMATION FOR SEQ ID NO: 610:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 34 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 610:

Met Val Leu Leu Leu Leu Leu Ala Tyr Val Leu Leu Thr Tyr Ile Leu  
1 5 10 15

Leu Leu Asn Met Leu Ile Ala Leu Met Xaa Arg Asp Arg Gln Gln Cys  
20 25 30

50 Arg His

(2) INFORMATION FOR SEQ ID NO: 611:

- (1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear



(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 611:

Met Val Phe Glu Gly Phe Ser Ala Phe Cys Leu Ser Thr Ala  
1 5 10 15  
Pro Thr Ser His Pro  
20

5

10

(2) INFORMATION FOR SEQ ID NO: 612:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 9 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 612:

Gly Lys Lys Asn Gln Leu Leu Val Ile  
1 5

20

(2) INFORMATION FOR SEQ ID NO: 613:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 29 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 613:

Met Val Trp Val Leu Trp Ser Ala Pro Ser Leu Ala Pro Pro Trp Val  
1 5 10 15

Gly Pro Cys Trp Pro Ser Thr Gly Asn Cys Cys Leu Cys  
20 25

35

(2) INFORMATION FOR SEQ ID NO: 614:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 30 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 614:

Met Ala Lys Arg Ser Pro Gly Gly Cys Gly Ser Gly Leu Ile Leu Leu  
1 5 10 15

Cys Cys Gln Pro Cys Arg Pro Thr Ser Ser Ala Pro Met Arg  
20 25 30

45

(2) INFORMATION FOR SEQ ID NO: 615:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 113 amino acids  
(B) TYPE: amino acid

60

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 615:

Ile Thr Ile Ala Ile Gln Met Ile Cys Leu Val Asn Xaa Glu Leu Tyr  
1 5 10 15

Pro Thr Phe Val Arg Asn Xaa Gly Val Met Val Cys Ser Ser Leu Cys  
20 25 30

Asp Ile Gly Gly Ile Ile Thr Pro Phe Ile Val Phe Arg Leu Arg Glu  
35 40 45

Val Trp Gln Ala Leu Pro Leu Ile Leu Phe Ala Val Leu Gly Leu Leu  
50 55 60

Ala Ala Gly Val Thr Leu Leu Leu Pro Gln Thr Lys Gly Val Ala Leu  
65 70 75 80

Pro Glu Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro  
85 90 95

Lys Glu Asn Thr Ile Tyr Leu Lys Val Gln Thr Ser Glu Pro Ser Gly  
100 105 110

25 Thr

(2) INFORMATION FOR SEQ ID NO: 616:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 18 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 616:

Thr Met Lys Asp Ala Glu Asn Leu Gly Arg Lys Ala Lys Pro Lys Glu  
1 5 10 15

40 Asn Thr

(2) INFORMATION FOR SEQ ID NO: 617:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 21 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 617:

Pro Arg Val Arg Asn Ser Pro Glu Asp Leu Gly Leu Ser Leu Thr Gly  
1 5 10 15

Asp Ser Cys Lys Leu  
20

60

(2) INFORMATION FOR SEQ ID NO: 618:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 52 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 618:

10 Gln Ala Asp Asp Leu Gln Ala Thr Val Ala Ala Leu Cys Val Leu Arg  
1 5 10 15

Gly Gly Gly Pro Trp Ala Gly Ser Trp Leu Ser Pro Lys Thr Thr Pro Gly  
20 25 30

15 Ala Met Gly Gly Asp Leu Val Leu Gly Leu Gly Ala Leu Arg Arg  
35 40 45

20 Lys Arg Leu Leu  
50

(2) INFORMATION FOR SEQ ID NO: 619:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 222 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 619:

30 Gln Gln Glu Lys Ser Leu Ala Gly Trp Ala Leu Val Leu Ala Xaa  
1 5 10 15

35 Gly Ile Gly Leu Met Val Leu His Ala Glu Met Leu Trp Phe Gly Gly  
20 25 30

Cys Ser Ala Val Asn Ala Thr Gly His Leu Ser Asp Thr Leu Trp Leu  
35 40 45

40 Ile Pro Ile Thr Phe Leu Thr Ile Gly Tyr Gly Asp Val Val Pro Gly  
50 55 60

45 Thr Met Trp Gly Lys Ile Val Cys Leu Cys Thr Gly Val Met Gly Val  
65 70 75 80

Cys Cys Thr Ala Leu Leu Val Ala Val Val Ala Arg Lys Leu Glu Phe  
85 90 95

50 Asn Lys Ala Glu Lys His Val His Asn Phe Met Met Asp Ile Gln Tyr  
100 105 110

Thr Lys Glu Met Lys Glu Ser Ala Ala Arg Val Leu Gln Glu Ala Trp  
115 120 125

55 Met Phe Tyr Lys His Thr Arg Arg Lys Glu Ser His Ala Ala Xaa  
130 135 140

His Gln Arg Xaa Leu Leu Ala Ile Asn Ala Phe Arg Gln Val Arg  
145 150 155 160

Leu Lys His Arg Lys Leu Arg Glu Gln Val Asn Ser Met Val Asp Ile  
165 170 175

5 Ser Lys Met His Met Ile Leu Tyr Asp Leu Gln Gln Asn Leu Ser Ser  
180 185 190

Ser His Arg Ala Leu Glu Lys Gln Ile Asp Thr Leu Ala Gly Lys Leu  
195 200 205

10 Asp Ala Leu Thr Glu Leu Leu Ser Thr Ala Leu Gly Pro Arg Gln Leu  
210 215 220

15 Pro Glu Pro Ser Gln Ser Lys  
225 230

(2) INFORMATION FOR SEQ ID NO: 620:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 36 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 620:

25 Tyr Gln Ala His His Val Ser Arg Asn Lys Arg Gly Gln Val Val Gly  
1 5 10 15

30 Thr Arg Gly Gly Phe Arg Gly Cys Thr Val Trp Leu Thr Gly Leu Ser  
20 25 30

Gly Ala Gly Lys  
35

(2) INFORMATION FOR SEQ ID NO: 621:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 57 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 621:

45 Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
1 5 10 15

50 Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
20 25 30

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
35 40 45

55 His Lys Ala Lys Ser His Pro Glu Val  
50 55

(2) INFORMATION FOR SEQ ID NO: 622:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 622:

Ile Thr Ser Thr Asp Ile Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln  
1 5 10 15

Pro Ser Asp

## (2) INFORMATION FOR SEQ ID NO: 623:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 623:

Asn Ser Thr Ser Gly Glu Cys Leu Leu Leu Glu Ala Glu Gly Met Ser  
1 5 10 15

Lys Ser Tyr

## (2) INFORMATION FOR SEQ ID NO: 624:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 51 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 624:

Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys Ile Phe  
1 5 10 15

Val Gly Ser Gly Ser Gly Thr Glu Gly Leu Val Met Asn Ser  
20 25 30

Asp Ile Leu Gly Ala Thr Thr Glu Val Ile Glu Asp Ser Asp Ser  
35 40 45

Ala Gly Pro

50

## (2) INFORMATION FOR SEQ ID NO: 625:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 625:

Asp Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
20 25 30

Ile Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala  
1 5 10 15

His Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys  
20 25 30

Lys Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu  
35 40 45

Gln Lys Gln Leu Leu Arg His Ala Lys His His Thr  
50 55 60

## (2) INFORMATION FOR SEQ ID NO: 626:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 31 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 626:

Asp Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala Arg Ala Phe Lys Ser  
1 5 10 15

Ser His Asn Leu Ala Val His Arg Met Ile His Thr Gly Glu Lys  
20 25 30

## (2) INFORMATION FOR SEQ ID NO: 627:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 627:

Arg Ser Ser Arg Ser Lys Thr Gly Ser Leu Gln Leu Ile Cys Lys Ser  
1 5 10 15

Glu Pro Asn Thr Asp Gln Leu Asp Tyr  
20 25

## (2) INFORMATION FOR SEQ ID NO: 628:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 183 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 628:

Leu Gln Cys Glu Ile Cys Gly Phe Thr Cys Arg Gln Lys Ala Ser Leu  
1 5 10 15

Asn Trp His Met Lys Lys His Asp Ala Asp Ser Phe Tyr Gln Phe Ser  
20 25 30

Cys Asn Ile Cys Gly Lys Lys Phe Glu Lys Lys Asp Ser Val Val Ala  
35 40 45

5 His Lys Ala Lys Ser His Pro Glu Val Xaa Ile Thr Ser Thr Asp Ile  
50 55 60

10 Leu Gly Thr Asn Pro Glu Ser Leu Thr Gln Pro Ser Asp Xaa Asn Ser  
65 70 75 80

Thr Ser Gly Glu Cys Leu Leu Glu Ala Glu Gly Met Ser Lys Ser  
85 90 95

15 Tyr Xaa Cys Ser Gly Thr Glu Arg Val Ser Leu Met Ala Asp Gly Lys  
100 105 110

Ile Phe Val Gly Ser Gly Ser Gly Gly Thr Glu Gly Leu Val Met  
115 120 125

20 Asn Ser Asp Ile Leu Gly Ala Thr Thr Glu Val Leu Ile Glu Asp Ser  
130 135 140

Asp Ser Ala Gly Pro Xaa Gln Arg Asp Tyr Ile Cys Glu Tyr Cys Ala  
145 150 155 160

25 Arg Ala Phe Lys Ser Ser His Asn Leu Ala Val His Arg Met Ile His  
165 170 175

30 Thr Gly Glu Lys His Tyr Xaa  
180

(2) INFORMATION FOR SEQ ID NO: 629:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 60 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 629:

Gln Tyr Val Arg Cys Glu Met Glu Gly Cys Gly Thr Val Leu Ala His  
1 5 10 15

45 Pro Arg Tyr Leu Gln His His Ile Lys Tyr Gln His Leu Leu Lys Lys  
20 25 30

Lys Tyr Val Cys Pro His Pro Ser Cys Gly Arg Leu Phe Arg Leu Gln  
35 40 45

50 Lys Gln Leu Leu Arg His Ala Lys His His Thr Asp  
50 55 60

(2) INFORMATION FOR SEQ ID NO: 630:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 630:

5 Pro Phe Lys Asp Asp Pro Arg Asp Glu Thr Tyr Lys Pro His Leu Glu  
1 5 10 15

10 Arg Glu Thr Pro Lys Pro Arg Arg Lys Ser Gly  
20 25

(2) INFORMATION FOR SEQ ID NO: 631:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 110 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 631:

20 Glu Met Phe Asp Ser Leu Ser Tyr Phe Lys Gly Ser Ser Leu Leu Leu  
1 5 10 15

25 Met Leu Lys Thr Tyr Leu Ser Glu Asp Val Phe Gln His Ala Val Val  
20 25 30

Leu Tyr Leu His Asn His Ser Tyr Ala Ser Ile Gln Ser Asp Asp Leu  
35 40 45

30 Trp Asp Ser Phe Asn Glu Val Thr Asn Gln Thr Leu Asp Val Lys Arg  
50 55 60

Met Met Lys Thr Trp Thr Leu Gln Lys Gly Phe Pro Leu Val Thr Val  
65 70 75 80

35 Gln Lys Lys Gly Lys Glu Leu Phe Ile Gln Gln Glu Arg Phe Phe Leu  
85 90 95

40 Asn Met Lys Pro Glu Ile Gln Pro Ser Asp Thr Arg Tyr Met  
100 105 110

(2) INFORMATION FOR SEQ ID NO: 632:

(1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(x1) SEQUENCE DESCRIPTION: SEQ ID NO: 632:

50 Leu Glu Lys Val Ala Ser Val Gly Asn Ser Arg Pro Thr Gly Gln Gln  
1 5 10 15

55 Leu Glu Ser Leu Gly Leu Leu Ala  
20

(2) INFORMATION FOR SEQ ID NO: 633:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 18 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 633:

Val His Arg Glu Glu Ala Ser Cys Tyr Cys Gln Ala Glu Pro Ser Gly  
1 5 10 15

10 Asp Leu

## (2) INFORMATION FOR SEQ ID NO: 634:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 92 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 634:

Arg Pro Ala Leu Arg Gln Ala Gly Gly Thr Arg Glu Pro Arg Gln  
1 5 10 15

Lys Arg Trp Ala Gly Leu  
20

## (2) INFORMATION FOR SEQ ID NO: 635:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 12 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 635:

Ala Val Asn Phe Arg Pro Gln Arg Ser Gln Ser Met  
1 5 10

## (2) INFORMATION FOR SEQ ID NO: 636:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 37 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 636:

Met Ile Thr Asp Val Gln Leu Ala Ile Phe Ala Asn Met Leu Gly Val  
1 5 10 15

Ser Leu Phe Leu Leu Val Val Leu Tyr His Tyr Val Ala Val Asn Asn  
20 25 30

Pro Lys Lys Gln Glu  
35

60

## (2) INFORMATION FOR SEQ ID NO: 637:

## (1) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 342 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 637:

Glu Glu Met Ala Asp Ser Val Lys Thr Phe Leu Gln Asp Leu Ala Arg  
1 5 10 15

Gly Ile Lys Asp Ser Ile Trp Gly Ile Cys Thr Ile Ser Lys Leu Asp  
20 25 30

Ala Arg Ile Gln Gln Lys Arg Glu Gln Arg Arg Arg Ala Ser  
35 40 45

Ser Val Leu Ala Gln Arg Arg Ala Gln Ser Ile Glu Arg Lys Gln Glu  
50 55 60

Ser Glu Pro Arg Ile Val Ser Arg Ile Phe Gln Cys Cys Ala Trp Asn  
65 70 75 80

Gly Gly Val Phe Trp Phe Ser Leu Leu Phe Tyr Arg Val Phe Ile  
85 90 95

Pro Val Leu Gln Ser Val Thr Ala Arg Ile Ile Gly Asp Pro Ser Leu  
100 105 110

His Gly Asp Val Trp Ser Trp Leu Glu Phe Phe Leu Thr Ser Ile Phe  
115 120 125

Ser Ala Leu Trp Val Leu Pro Leu Phe Val Leu Ser Lys Val Val Asn  
130 135 140

Ala Ile Trp Phe Gln Asp Ile Ala Asp Leu Ala Phe Glu Val Ser Gly  
145 150 155 160

Arg Lys Pro His Pro Phe Pro Ser Val Ser Lys Ile Ile Ala Asp Met  
165 170 175

Leu Phe Asn Leu Leu Leu Gln Ala Leu Phe Leu Ile Gln Gly Met Phe  
180 185 190

Val Ser Leu Phe Pro Ile His Leu Val Gly Gln Leu Val Ser Leu Leu  
195 200 205

His Met Ser Leu Leu Tyr Ser Leu Tyr Cys Phe Glu Tyr Arg Trp Phe  
210 215 220

Asn Lys Gly Ile Glu Met His Gln Arg Leu Ser Asn Ile Glu Arg Asn  
225 230 235 240

Trp Pro Tyr Tyr Phe Gly Phe Gly Leu Pro Leu Ala Phe Leu Thr Ala  
245 250 255

Met Gln Ser Ser Tyr Ile Ile Ser Gly Cys Leu Phe Ser Ile Leu Phe  
260 265 270

Pro Leu Phe Ile Ile Ser Ala Asn Glu Ala Lys Thr Pro Gly Lys Ala  
275 280 285

5 Tyr Leu Phe Gln Arg Leu Phe Ser Leu Val Val Phe Leu Ser Asn  
290 295 300

Arg Leu Phe His Lys Thr Val Tyr Leu Gln Ser Ala Leu Ser Ser Ser  
305 310 315 320

10 Thr Ser Ala Glu Lys Phe Pro Ser Pro His Pro Ser Pro Ala Lys Leu  
325 330 335

15 Lys Ala Thr Ala Gly His  
340

(2) INFORMATION FOR SEQ ID NO: 638:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 529 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 638:

Met Ala Lys Phe Met Thr Pro Val Ile Gln Asp Asn Pro Ser Gly Trp  
1 5 10 15

30 Gly Pro Cys Ala Val Pro Glu Gln Phe Arg Asp Met Pro Tyr Gln Pro  
20 25 30

Phe Ser Lys Gly Asp Arg Leu Gly Lys Val Ala Asp Trp Thr Gly Ala  
35 40 45

35 Thr Tyr Gln Asp Lys Arg Tyr Thr Asn Lys Tyr Ser Ser Gln Phe Gly  
50 55 60

Gly Gly Ser Gln Tyr Ala Tyr Phe His Glu Glu Asp Glu Ser Ser Phe  
65 70 75 80

Gln Leu Val Asp Thr Ala Arg Thr Gln Lys Thr Ala Tyr Gln Arg Asn  
85 90 95

45 Arg Met Arg Phe Ala Gln Arg Asn Leu Arg Arg Asp Lys Asp Arg Arg  
100 105 110

Asn Met Leu Gln Phe Asn Leu Gln Ile Leu Pro Lys Ser Ala Lys Gln  
115 120 125

50 Lys Glu Arg Glu Arg Ile Arg Leu Gln Lys Lys Phe Gln Lys Gln Phe  
130 135 140

Gly Val Arg Gln Lys Trp Asp Gln Lys Ser Gln Lys Pro Arg Asp Ser  
145 150 155 160

Ser Val Glu Val Arg Ser Asp Trp Glu Val Lys Glu Glu Met Asp Phe  
165 170 175

60 Pro Gln Leu Met Lys Met Arg Tyr Tyr Leu Glu Val Ser Glu Pro Gln Asp

180 185 190

Ile Glu Cys Cys Gly Ala Leu Glu Tyr Tyr Asp Lys Ala Phe Asp Arg  
195 200 205

5 Ile Thr Thr Arg Ser Glu Lys Pro Leu Arg Xaa Lys Arg Ile Phe  
210 215 220

His Thr Val Thr Thr Asp Asp Pro Val Ile Arg Lys Leu Ala Lys  
225 230 235 240

10 Thr Gln Gly Asn Val Phe Ala Thr Asp Ala Ile Leu Ala Thr Leu Met  
245 250 255

15 Ser Cys Thr Arg Ser Val Tyr Ser Trp Asp Ile Val Val Gln Arg Val  
260 265 270

Gly Ser Lys Leu Phe Phe Asp Lys Arg Asp Asn Ser Asp Phe Asp Leu  
275 280 285

20 Leu Thr Val Ser Glu Thr Ala Asn Glu Pro Pro Gln Asp Glu Gly Asn  
290 295 300

Ser Phe Asn Ser Pro Arg Asn Leu Ala Met Glu Ala Thr Tyr Ile Asn  
305 310 315 320

25 His Asn Phe Ser Gln Gln Cys Leu Arg Met Gly Lys Glu Arg Tyr Asn  
325 330 335

30 Phe Pro Asn Pro Asn Pro Phe Val Glu Asp Asp Met Asp Lys Asn Glu  
340 345 350

Ile Ala Ser Val Ala Tyr Arg Tyr Arg Ser Gly Lys Leu Gly Asp Asp  
355 360 365

35 Ile Asp Leu Ile Val Arg Cys Glu His Asp Gly Val Met Thr Gly Ala  
370 375 380

Asn Gly Glu Val Ser Phe Ile Asn Ile Lys Thr Thr Leu Asn Glu Trp Asp  
385 390 395 400

40 Ser Arg His Cys Asn Gly Val Asp Trp Arg Gln Lys Leu Asp Ser Gln  
405 410 415

45 Arg Gly Ala Val Ile Ala Thr Glu Leu Lys Asn Asn Ser Tyr Lys Leu  
420 425 430

Ala Arg Trp Thr Cys Cys Ala Leu Leu Ala Gly Ser Glu Tyr Leu Lys  
435 440 445

50 Leu Gly Tyr Val Ser Arg Tyr His Val Lys Asp Ser Ser Arg His Val  
450 455 460

Ile Leu Gly Thr Gln Gln Phe Lys Pro Asn Glu Phe Ala Ser Gln Ile  
465 470 475 480

55 Asn Leu Ser Val Glu Asn Ala Trp Gly Ile Leu Arg Cys Val Ile Asp  
485 490 495

60 Ile Cys Met Lys Leu Glu Glu Gly Lys Tyr Leu Ile Leu Lys Asp Pro

500 505 510  
 Asn Lys Gln Val Ile Arg Val Tyr Ser Leu Pro Asp Gly Thr Phe Ser  
 515 520 525  
 5 Ser

10 (2) INFORMATION FOR SEQ ID NO: 639:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 194 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 639:  
 Lys Lys Arg His Thr Asp Val Gln Phe Tyr Thr Glu Val Gly Glu Ile  
 1 5 10 15  
 Thr Thr Asp Leu Gly Lys His Gln His Met His Asp Arg Asp Asp Leu  
 20 25 30  
 Tyr Ala Glu Gln Met Glu Arg Glu Met Arg His Lys Leu Lys Thr Ala  
 35 40 45  
 Phe Lys Asn Phe Ile Glu Lys Val Glu Ala Leu Thr Lys Glu Glu Leu  
 50 55 60  
 Glu Phe Glu Val Pro Phe Arg Asp Leu Gly Phe Asn Gly Ala Pro Tyr  
 65 70 75 80  
 Arg Ser Thr Cys Leu Leu Gln Pro Thr Ser Ser Ala Leu Val Asn Ala  
 85 90 95  
 Thr Glu Trp Pro Pro Phe Val Thr Leu Asp Glu Val Glu Leu Ile  
 100 105 110  
 His Phe Xaa Arg Val Gln Phe His Leu Lys Asn Phe Asp Met Val Ile  
 115 120 125  
 Val Tyr Lys Asp Tyr Ser Lys Lys Val Thr Met Ile Asn Ala Ile Pro  
 130 135 140  
 Val Ala Ser Leu Asp Pro Ile Lys Glu Trp Leu Asn Ser Cys Asp Leu  
 145 150 155 160  
 Lys Tyr Thr Glu Gly Val Gln Ser Leu Asn Trp Thr Lys Ile Met Lys  
 165 170 175  
 Thr Ile Val Asp Asp Pro Glu Gly Phe Phe Glu Gln Gly Gly Trp Ser  
 180 185 190  
 55 Phe Leu

60 (2) INFORMATION FOR SEQ ID NO: 640:

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 233 amino acids  
 (B) TYPE: amino acid

(1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 70 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 640:  
 5 Arg Ser Gly Leu Gly Leu Gly Ile Thr Ile Ala Phe Leu Ala Thr Leu  
 1 5 10 15  
 10 Ile Thr Gln Phe Leu Val Tyr Asn Gly Val Tyr Gln Tyr Thr Ser Pro  
 20 25 30  
 Asp Phe Leu Tyr Ile Arg Ser Trp Leu Pro Cys Ile Phe Phe Ser Gly  
 35 40 45  
 Gly Val Thr Val Gly Asn Ile Gly Arg Gln Leu Ala Met Gly Val Pro  
 50 55 60  
 20 Glu Lys Pro His Ser Asp  
 65 70

25 (2) INFORMATION FOR SEQ ID NO: 641:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 101 amino acids  
 (B) TYPE: amino acid  
 (D) TOPOLOGY: linear  
 (xl) SEQUENCE DESCRIPTION: SEQ ID NO: 641:  
 30 Val Thr Gln Pro Lys His Leu Ser Ala Ser Met Gly Gly Ser Val Glu  
 1 5 10 15  
 35 Ile Pro Phe Ser Phe Tyr Tyr Pro Trp Glu Leu Ala Xaa Pro Xaa  
 20 25 30  
 Val Arg Ile Ser Trp Arg Arg Gly His Phe His Gly Gln Ser Phe Tyr  
 35 40 45  
 Ser Thr Arg Pro Pro Ser Ile His Lys Asp Tyr Val Asn Arg Leu Phe  
 50 55 60  
 45 Leu Asn Trp Thr Glu Gly Gln Glu Ser Gly Phe Leu Arg Ile Ser Asn  
 65 70 75 80  
 Leu Arg Lys Glu Asp Gln Ser Val Tyr Phe Cys Arg Val Glu Leu Asp  
 85 90 95  
 50 Thr Arg Arg Ser Gly  
 100

55 (2) INFORMATION FOR SEQ ID NO: 642:  
 (1) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 233 amino acids  
 (B) TYPE: amino acid

718

(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 642:

5 Met Glu Ala Gln Gln Val Asn Glu Ala Glu Ser Ala Arg Glu Gln Leu 15  
1 5 10  
Gln Xaa Leu His Asp Gln Ile Ala Gly Gln Lys Ala Ser Lys Gln Glu 30  
20 25  
10 Leu Glu Thr Glu Leu Glu Arg Leu Lys Gln Glu Phe His Tyr Ile Glu 45  
35  
Glu Asp Leu Tyr Arg Thr Lys Asn Thr Leu Gln Ser Arg Ile Lys Asp 60  
50 55  
15 Arg Asp Glu Glu Ile Gln Lys Leu Arg Asn Gln Leu Thr Asn Lys Thr 75  
65 80  
Leu Ser Asn Ser Ser Gln Ser Glu Leu Glu Asn Arg Leu His Gln Leu 95  
85 90  
20 Thr Glu Thr Leu Ile Gln Lys Gln Thr Met Leu Glu Ser Leu Ser Thr 110  
100 105  
25 Glu Lys Asn Ser Leu Val Phe Gln Leu Glu Arg Leu Glu Gln Met 125  
115 120  
Asn Ser Ala Ser Gly Ser Ser Asn Gly Ser Ser Ile Asn Met Ser 140  
130 135  
30 Gly Ile Asp Asn Gly Glu Gly Thr Arg Leu Arg Asn Val Pro Val Leu 150  
145 155  
Phe Asn Asp Thr Glu Thr Asn Leu Ala Gly Met Tyr Gly Lys Val Arg 175  
165 170  
35 Lys Ala Ala Ser Ser Ile Asp Gln Phe Ser Ile Arg Leu Gly Ile Phe 185  
180  
Leu Arg Arg Tyr Pro Ile Ala Arg Val Phe Val Ile Ile Tyr Met Ala 205  
195 200  
40 Leu Leu His Leu Trp Val Met Ile Val Leu Leu Thr Tyr Thr Pro Glu 220  
210 215  
45 Met His His Asp Gln Pro Tyr Gly Lys 230  
225 230

(2) INFORMATION FOR SEQ ID NO: 643:

(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 43 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 643:

50  
55  
60 Ile Arg His Glu Gln His Pro Asn Phe Ser Leu Glu Met His Ser Lys 15  
1 5 10

719

5 Val Cys Ala His Leu His Glu Glu Leu Asn Cys 40  
35  
10 (2) INFORMATION FOR SEQ ID NO: 644:  
(1) SEQUENCE CHARACTERISTICS:  
(A) LENGTH: 63 amino acids  
(B) TYPE: amino acid  
(D) TOPOLOGY: linear  
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 644:  
Ser Phe Phe Ile Ser Glu Glu Lys Gly His Leu Leu Leu Gln Ala Glu 15  
1 5 10  
20 Arg His Pro Trp Val Ala Gly Ala Leu Val Gly Val Ser Gly Gly Leu 25  
30  
Thr Leu Thr Thr Cys Ser Gly Pro Thr Glu Lys Pro Ala Thr Lys Asn 45  
35 40  
Tyr Phe Leu Lys Arg Leu Leu Gln Glu Met His Ile Arg Ala Asn 60  
50 55



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97897
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 116, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209043
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 119, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit September 4, 1997	Accession Number 209235
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
Further deposits are identified on an additional sheet <input type="checkbox"/>	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97898
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 122, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution: American Type Culture Collection Address of depository institution (including postal code and country): 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit: May 15, 1997 Accession Number: 209044 C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer: Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer:	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution: American Type Culture Collection Address of depository institution (including postal code and country): 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit: February 26, 1997 Accession Number: 97899 C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer: Susan White PCT International Division	
For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer:	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 126, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit May 15, 1997 Accession Number 209045 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White POT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 130, line N/A	
B. IDENTIFICATION OF DEPOSIT Name of depository institution American Type Culture Collection Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America Date of deposit April 28, 1997 Accession Number 209011 C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only <input checked="" type="checkbox"/> This sheet was received with the international application Authorized officer Susan White PCT International Division	For International Bureau use only <input type="checkbox"/> This sheet was received by the International Bureau on: Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97900
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97901
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	
Susan White PCT International Division	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 131, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209046
C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 136*is*)

A. The indications made below relate to the microorganism referred to in the description on page 137, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	May 15, 1997
Accession Number	209047
C. ADDITIONAL INDICATIONS (new blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (new blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only	
This sheet was received with the international application <input checked="" type="checkbox"/>	
Authorized officer Susan White PCT International Division	
For International Bureau use only	
This sheet was received by the International Bureau on: <input type="checkbox"/>	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 137 . line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 22, 1997	Accession Number 209076
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 140 . line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit August 21, 1997	Accession Number 209215
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97904
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 134, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit July 3, 1997	Accession Number 209139
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (If the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer



## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 153, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209049
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 153, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depositary institution American Type Culture Collection	
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit February 26, 1997	Accession Number 97903
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution	American Type Culture Collection
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	June 12, 1997
Accession Number	209119
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depositary institution	American Type Culture Collection
Address of depositary institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 26, 1997
Accession Number	97902
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 146, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209048
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 160, line N/A	
B. IDENTIFICATION OF DEPOSIT Further deposits are identified on an additional sheet <input type="checkbox"/>	
Name of depository institution American Type Culture Collection	
Address of depository institution (including postal code and country) 12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit May 15, 1997	Accession Number 209050
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable) The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only For International Bureau use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	<input type="checkbox"/> This sheet was received by the International Bureau on:
Authorized officer Susan White PCT International Division	Authorized officer

## INDICATIONS RELATING TO A DEPOSITED MICROORGANISM

(PCT Rule 13bis)

A. The indications made below relate to the microorganism referred to in the description on page 142, line N/A	
B. IDENTIFICATION OF DEPOSIT	
Name of depository institution	American Type Culture Collection
Address of depository institution (including postal code and country)	
12301 Parklawn Drive Rockville, Maryland 20852 United States of America	
Date of deposit	February 12, 1998
Accession Number	209627
C. ADDITIONAL INDICATIONS (leave blank if not applicable) This information is continued on an additional sheet <input type="checkbox"/>	
D. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE (if the indications are not for all designated States)	
E. SEPARATE FURNISHING OF INDICATIONS (leave blank if not applicable)	
The indications listed below will be submitted to the International Bureau later (specify the general nature of the indications, e.g., "Accession Number of Deposit")	
For receiving Office use only	
<input checked="" type="checkbox"/> This sheet was received with the international application	
Authorized officer	
Susan White PCT International Division	
For International Bureau use only	
<input type="checkbox"/> This sheet was received by the International Bureau on:	
Authorized officer	

## What Is Claimed Is:

1. An isolated nucleic acid molecule comprising a polynucleotide having a nucleotide sequence at least 95% identical to a sequence selected from the group consisting of:
  - (a) a polynucleotide fragment of SEQ ID NO:X or a polynucleotide fragment of the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (b) a polynucleotide encoding a polypeptide fragment of SEQ ID NO:Y or a polypeptide fragment encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (c) a polynucleotide encoding a polypeptide domain of SEQ ID NO:Y or a polypeptide domain encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (d) a polynucleotide encoding a polypeptide epitope of SEQ ID NO:Y or a polypeptide epitope encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X;
  - (e) a polynucleotide encoding a polypeptide of SEQ ID NO:Y or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X, having biological activity;
  - (f) a polynucleotide which is a variant of SEQ ID NO:X;
  - (g) a polynucleotide which is an allelic variant of SEQ ID NO:X;
  - (h) a polynucleotide which encodes a species homologue of the SEQ ID NO:Y;
  - (i) a polynucleotide capable of hybridizing under stringent conditions to any one of the polynucleotides specified in (a)-(h), wherein said polynucleotide does not hybridize under stringent conditions to a nucleic acid molecule having a nucleotide sequence of only A residues or of only T residues.
2. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding a secreted protein.
3. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises a nucleotide sequence encoding the sequence identified as SEQ ID NO:Y or the polypeptide encoded by the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

4. The isolated nucleic acid molecule of claim 1, wherein the polynucleotide fragment comprises the entire nucleotide sequence of SEQ ID NO:X or the cDNA sequence included in ATCC Deposit No:Z, which is hybridizable to SEQ ID NO:X.

5

5. The isolated nucleic acid molecule of claim 2, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

10

6. The isolated nucleic acid molecule of claim 3, wherein the nucleotide sequence comprises sequential nucleotide deletions from either the C-terminus or the N-terminus.

15

7. A recombinant vector comprising the isolated nucleic acid molecule of claim 1.

20

8. A method of making a recombinant host cell comprising the isolated nucleic acid molecule of claim 1.

25

9. A recombinant host cell produced by the method of claim 8.

30

10. The recombinant host cell of claim 9 comprising vector sequences.

35

11. An isolated polypeptide comprising an amino acid sequence at least 95% identical to a sequence selected from the group consisting of:

(a) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(b) a polypeptide fragment of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z, having biological activity;

(c) a polypeptide domain of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(d) a polypeptide epitope of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(e) a secreted form of SEQ ID NO:Y or the encoded sequence included in

40

ATCC Deposit No:Z;

(f) a full length protein of SEQ ID NO:Y or the encoded sequence included in ATCC Deposit No:Z;

(g) a variant of SEQ ID NO:Y;

(h) an allelic variant of SEQ ID NO:Y; or

(i) a species homologue of the SEQ ID NO:Y.

12. The isolated polypeptide of claim 11, wherein the secreted form or the full length protein comprises sequential amino acid deletions from either the C-terminus or the N-terminus.

45

13. An isolated antibody that binds specifically to the isolated polypeptide of claim 11.

50

14. A recombinant host cell that expresses the isolated polypeptide of claim 11.

55

15. A method of making an isolated polypeptide comprising:

(a) culturing the recombinant host cell of claim 14 under conditions such that said polypeptide is expressed; and

(b) recovering said polypeptide.

60

16. The polypeptide produced by claim 15.

65

17. A method for preventing, treating, or ameliorating a medical condition, comprising administering to a mammalian subject a therapeutically effective amount of the polypeptide of claim 11 or the polynucleotide of claim 1.

70

18. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or absence of a mutation in the polynucleotide of claim 1; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or absence of said mutation.

75

19. A method of diagnosing a pathological condition or a susceptibility to a pathological condition in a subject comprising:

(a) determining the presence or amount of expression of the polypeptide of claim 11 in a biological sample; and

(b) diagnosing a pathological condition or a susceptibility to a pathological condition based on the presence or amount of expression of the polypeptide.

20. A method for identifying a binding partner to the polypeptide of claim 11 comprising:
- 5 (a) contacting the polypeptide of claim 11 with a binding partner; and  
(b) determining whether the binding partner effects an activity of the polypeptide.
21. The gene corresponding to the cDNA sequence of SEQ ID NO:Y.
- 10 22. A method of identifying an activity in a biological assay, wherein the method comprises:
- 15 (a) expressing SEQ ID NO:X in a cell;  
(b) isolating the supernatant;  
(c) detecting an activity in a biological assay; and  
(d) identifying the protein in the supernatant having the activity.
23. The product produced by the method of claim 22.

## KiSS-1, a Novel Human Malignant Melanoma Metastasis-Suppressor Gene

Jeong-Hyung Lee, Mary E. Miele, Deana J. Hicks, Karen K. Phillips, Jeffery M. Trent, Bernard E. Weissman, Danny R. Welch\*

**Background:** Microcell-mediated transfer of chromosome 6 into human C8161 and MelJuSo melanoma cells suppresses their ability to metastasize by at least 95% without affecting their tumorigenicity. This observation demonstrates that the ability to metastasize is a phenotype distinct from tumor formation and suggests that tumorigenic cells acquire metastatic capability only after accumulating additional genetic defects. These results also imply that mutations of genes on chromosome 6 are among those late genetic changes responsible for metastatic potential. They further suggest that a melanoma metastasis-suppressor gene(s) is encoded on chromosome 6 or is regulated by genes on chromosome 6. **Purpose:** Our objective was to identify the gene(s) responsible for the suppression of metastasis in chromosome 6/melanoma cell hybrids. **Methods:** A modified subtractive hybridization technique was used to compare the expression of messenger RNAs (mRNAs), via an analysis of complementary DNAs (cDNAs), in metastatic cells (C8161 or MelJuSo) and nonmetastatic hybrid clones (neo6/C8161 or neo6/MelJuSo). **Results:** A novel cDNA, designated KiSS-1, was isolated from malignant melanoma cells that had been suppressed for metastatic potential by the introduction of human chromosome 6. Northern blot analyses comparing mRNAs from a panel of human melanoma cells revealed that KiSS-1 mRNA expression occurred only in nonmetastatic melanoma cells. Expression of this mRNA in normal heart, brain, liver, lung, and skeletal muscle was undetectable by northern blot analysis. Weak expression was found in the kidney and pancreas, but the highest expression was observed in the placenta. The KiSS-1 cDNA encodes a predominantly hydrophilic, 164 amino acid protein with a polyproline-rich domain indicative of an SH3 ligand (binds to the homology 3 domain of the oncoprotein Src) and a putative protein kinase C- $\alpha$  phosphorylation site. Transfection of a full-length KiSS-1 cDNA into C8161 melanoma cells suppressed metastasis in an expression-dependent manner. **Conclusions:** These data strongly suggest that KiSS-1 expression may suppress the metastatic potential of malignant melanoma cells. **Implications:** KiSS-1 may be a useful marker for distinguishing metastatic melanomas from

nonmetastatic melanomas. [J Natl Cancer Inst 1996;88:1731-7]

The spread of malignant tumor cells from a primary tumor to form metastases at distant sites is the most life-threatening complication of cancer and is responsible for the majority of deaths in affected individuals. Understanding the regulation of metastasis at the molecular level is required to devise new modalities of cancer therapy, particularly those that improve cure rates when treating metastatic cancer.

Metastasis is a multistep process involving complex interactions between tumor cells and host cells. To metastasize, tumor cells must invade from the primary tumor, dissociate from the tumor mass, and be transported to nearby or distant secondary sites. Single cells, homotypic clusters of cells, or heterotypic emboli then arrest at a distant site with the use of both organ-specific and nonspecific mechanisms, invade into the surrounding tissue, and respond to growth signals at the secondary site (1-4). A tumor cell must successfully accomplish each step in the pathway or metastases will not develop. Both positive and negative regulators exist for each step in the metastatic cascade (5-11), implicating the involvement of dozens of different genes. This provides the challenging task of identifying critical genes controlling the process for use as targets for therapeutic intervention.

To identify metastasis-controlling genes, we have used a functional assay of metastasis suppression of human tumor cells in athymic nude mice. We previously showed that the introduction of human chromosome 6 into human metastatic melanoma

\*Affiliations of authors: J.-H. Lee, M. E. Miele, D. J. Hicks, D. R. Welch, Department of Experimental Pathology, The Jake Gittlen Cancer Research Institute, College of Medicine, The Pennsylvania State University, Hershey; K. K. Phillips, B.E. Weissman, Department of Pathology, Lineberger Cancer Center, University of North Carolina, Chapel Hill; J. M. Trent, Division of Intramural Research, National Center for Human Genome Research, Bethesda, MD.

Correspondence to: Danny R. Welch, Ph.D., Department of Experimental Pathology (C7810), College of Medicine, The Pennsylvania State University, 500 University Dr., Hershey, PA 17033-0850.

See "Notes" section following "References."

cell lines C8161 or MelJuSo by microcell-mediated chromosome transfer resulted in at least 95% suppression of metastasis without affecting tumorigenicity (12,13) or local invasiveness (14). Introduction of chromosome 1 also partially suppressed MelJuSo metastasis (13). The development of this panel of cell lines provided us with the necessary reagents to isolate genes specifically involved in the control of metastasis.

To identify the gene(s) responsible for metastasis suppression in malignant melanoma, we used a modified subtractive hybridization method (15-19). Candidate genes were identified by searching for a minimum 10-fold increase in messenger RNA (mRNA) expression in nonmetastatic chromosome 6-C8161 hybrid cells (neo6/C8161.1) compared with highly metastatic parental C8161 cells grown under similar conditions.

## Materials and Methods

### Cell Lines and Culture Conditions

C8161 is an amelanotic human melanoma cell line that metastasizes widely following intradermal, subcutaneous, or intravenous injection into athymic nude or severe combined immunodeficient mice (12,20). Subclones C8161cl.9 and C8161cl.8 were isolated from C8161 by limiting dilution and were chosen for these studies because they have the highest and lowest metastatic potentials, respectively (12). neo6/C8161.1, neo6/C8161.2, and neo6/C8161.3 (12) were derived from microcell-mediated chromosome transfer of a single copy of a neomycin-resistance gene-tagged human chromosome 6 with the use of the MCH262A1.D6 donor cell line (12,21). Four additional, independently derived neo6/C8161 hybrid cell clones (neo6/C8161.4, neo6/C8161.5b, neo6/C8161.6, and neo6/C8161.8) were subsequently developed. None of the neo6/C8161 hybrid cell clones were metastatic in athymic nude mice. neo6(del)(q21-q23) hybrid cell clones were prepared using a related chromosome 6 microcell donor variant A9/6q-cl.1.2 (22) and are metastatic in nude mice (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR: manuscript in preparation).

Cell line nomenclature has been chosen to identify the origin and nature of each cell line as unambiguously as possible. Single-cell clones are preceded by a "c" (e.g., C8161cl.9 is a single-cell clone), and pooled, uncloned populations are identified by a "p" (e.g., C8161-KiSS-1-P1 is an uncloned population No. 1 of C8161 cells transfected with KiSS-1). Microcell hybrids are identified by the tagged chromosome followed by a "c" [e.g., neo6/C8161.3 is single-cell clone 3 derived from a fusion with a neomycin-resistance gene-tagged human chromosome 6; neo6(del)(q21-q23)/C8161cl.8 is single-cell clone 8 derived from a microcell fusion with a neotagged chromosome 6 containing deletion of the q21-q23 bands].

All cells were grown in Dulbecco's modified minimal Eagle's medium and Ham's F-12 medium (DME-F12) (Irvine Scientific Co., Santa Ana, CA) supplemented with 10% fetal bovine serum (cDME-F12). Neo-resistant cells, neo6/C8161, and neo6(del)(q21-q23)/C8161 hybrids were maintained in cDME-F12 containing 500 µg/mL geneticin (G-418; Life Technologies, Inc. [GIBCO BRL], Gaithersburg, MD). Nearly 70%-90% confluent cultures were passaged in Corning (Oneonta, NY) tissue culture dishes following detachment with a solution of calcium- and magnesium-free Dulbecco's phosphate-buffered saline containing 2 mM EDTA at split ratios of 1:10 to 1:20. All cultures were routinely tested and found to be negative for *Mycoplasma* spp. infection using a polymerase chain reaction (PCR)-based test kit (Pan Vera, Madison, WI).

### Subtractive Hybridization

Subtractive hybridization was done as described (18,19,23) with minor modifications. Briefly, mRNAs were isolated from 80% to 90% confluent cells using the FastTrack mRNA isolation kit (Invitrogen Corp., San Diego, CA). A complementary DNA (cDNA) library was constructed from neo6/C8161.1 cells using a λZAP-cDNA Gigapack II Gold cloning kit (Stratagene, La Jolla, CA). Subtractive hybridization was carried out between the first-strand cDNA from nonmetastatic neo6/C8161.1 cells and mRNA from metastatic parental C8161 cells. cDNA was synthesized from 1 µg polyadenylic acid [poly(A)<sup>+</sup>]-enriched RNA from neo6/C8161.1 using oligo(dT) primers. Biotinylation of mRNA from

parental C8161 cells was performed using the Photobiotin-labeling system (Life Technologies, Inc.) according to manufacturer's instructions. To subtract, biotinylated mRNA (20 µg) was mixed with cDNA (1 µg) from neo6/C8161.1 and ethanol precipitated. The pellet was resuspended in diethylpyrocyanate-treated water (20.5 µL), then added to 2× hybridization buffer (80% formamide, 100 mM hydroxyethylpiperazine ethane sulfonic acid [pH 7.5], and 2% sodium dodecyl sulfate [SDS], 22.5 µL). This mixture was boiled for 2 minutes, chilled on ice for 5 minutes, and followed by the addition of NaCl (5 M, 2 µL). Hybridization was performed at 42 °C for 48 hours. Streptavidin (25 µg; Sigma Chemical Co., St. Louis, MO) was added to the hybridization mixture, and the mixture was incubated at room temperature for 5 minutes before extraction with phenol-chloroform-isoamyl alcohol (25:24:1). The organic phase was back-extracted twice with 50 µL streptavidin-binding buffer (100 mM Tris-HCl [pH 8.0], 1 mM EDTA, and 500 mM NaCl). The aqueous phases were then pooled. Unhybridized cDNA target was ethanol precipitated. This subtracted cDNA was used as a probe for the neo6/C8161.1 cDNA library screening following random primer labeling (Amersham Life Sciences, Inc., Arlington Heights, IL). Positive clones were isolated and used for evaluation of expression in northern blot analysis.

### Screening Candidate cDNAs and Isolation of Full-Length KiSS-1

Candidate cDNAs were initially evaluated for differential expression by northern blot analysis using poly(A)<sup>+</sup>-enriched mRNAs from C8161 and neo6/C8161.1 cells. If the expression was greater in neo6/C8161.1 cells by at least 10-fold, the same cDNA was used to probe a more extensive blot containing RNAs from a panel of cell lines with differing metastatic potentials. If the pattern of expression correlated with metastatic potential, full-length cDNAs were isolated and further characterized.

Full-length KiSS-1 was obtained from a λZAP-cDNA library (7.5 × 10<sup>6</sup> plaques) prepared from neo6/C8161.1 cells. After third-round screening, seven positive clones were isolated and identified as a same gene by restriction mapping and sequence analysis. Library screening was done using a 0.5-kilobase (kb) partial KiSS-1 probe obtained from the subtractive hybridization. The longest cDNA insert was sequenced on both strands by the dideoxy chain termination method with Sequenase version 2.0 (Amersham Life Sciences, Inc.).

### Northern Blot Analysis

For northern blot analysis, poly(A)<sup>+</sup>-enriched mRNA was isolated from 80% to 90% confluent cells with the use of a FastTrack mRNA isolation kit (Invitrogen). Poly(A)<sup>+</sup> mRNA (2.5 µg) was loaded and electrophoresed on 1% agarose gel containing 2.2 M formaldehyde at 78 V for 3.5 hours. RNA was transferred onto a nylon membrane using the Turboblott system (Schleicher & Schuell, Keene, NH) and fixed by UV cross-linking (Stratagene). Full-length KiSS-1 cDNA probe was radiolabeled using random priming (Amersham Life Sciences, Inc.). Prehybridization was done in 48% formamide, 4.8× sodium chloride/sodium citrate (SSC) (1× SSC = 0.15 M NaCl and 0.015 M sodium citrate), 20 mM Tris (pH 7.6), 1× Denhardt's solution (0.02% Ficoll, 0.02% bovine serum albumin, and 0.02% polyvinylpyrrolidone), 10% dextran sulfate, and 0.1% SDS. Hybridization was carried out in the same buffer plus herring sperm DNA (100 µg/mL) for 24 hours at 42 °C. The membrane was rinsed twice with 2× SSC containing 0.1% SDS for 15 minutes at 42 °C, followed by washing twice with 0.1× SSC containing 0.1% SDS for 15 minutes at 42 °C. The rinsed membrane was exposed to x-ray film (Du Pont-NEN, Boston, MA) for desired intensity.

Expression of KiSS-1 in normal human tissues was evaluated by use of a multiple-tissue RNA blot (2 µg of poly(A) RNA per lane) purchased from CLONTECH Laboratories, Inc. (Palo Alto, CA). The blot was hybridized by use of a full-length KiSS-1 probe as above. Expression of KiSS-1 in normal human melanocytes (CLONTECH Laboratories, Inc.) was determined by reverse transcription-polymerase chain reaction (RT-PCR). Detection of KiSS-1 message (1.0 kb) in placenta was possible following overnight exposure. KiSS-1 mRNA was detectable in the pancreas (0.8 kb) and kidney (0.9 kb) only after a 3-7-day exposure at -70 °C with two intensifying screens.

### Transfections and Metastasis Assays

A full length of KiSS-1 cDNA was cloned into the pcDNA3 expression vector (Invitrogen Corp.) in which transcription is driven by human cytomegalovirus promoter. The resultant plasmid pcDNA3-KiSS-1 was transfected



into C8161 cells using Lipofectin (Life Technologies, Inc.). Individual transfectants were isolated and cloned following growth in cDME-F12 containing G-418 (500 µg/mL). Poly(A)<sup>+</sup> mRNA was isolated from transfectant by use of the Micro-FastTrack mRNA isolation kit (Invitrogen Corp.), and northern blot analysis was done as described above. C8161-neo6 mix was isolated as a control, uncloned population of C8161 cells transfected with the pcDNA3 vector only.

Transfectants were evaluated for growth in vitro. Cells ( $2 \times 10^4$ ) were inoculated into 24-well tissue culture plates (Corning) in cDME-F12 medium. After 24, 48, 72, and 96 hours, cells were trypsinized and counted using a hemacytometer.

For spontaneous metastasis assays, cells ( $1 \times 10^6$ ) were injected intradermally into the dorsolateral flank of athymic nude mice (four to 12 mice per group). Female mice (3–4 weeks old) (Harlan Sprague-Dawley, Inc., Madison, WI) were used for these studies. Animals were maintained under the guidelines of the National Institutes of Health and The Pennsylvania State University College of Medicine. All protocols were approved by the Institutional Animal Care and Use Committee. Food and water were provided ad libitum. Tumor size was measured weekly by taking orthogonal measurements and were expressed as mean tumor diameter. Mean tumor diameter was calculated as described (12,13) by use of the following equation:

$$\sqrt{(\text{diameter}_x) \times (\text{diameter}_y)}$$

After the mean tumor diameter reached 1.5–2.0 cm, mice were killed, necropsied, and visible metastases were counted (20,24).

In the C8161 human melanoma model, the rank orders of spontaneous and experimental metastasis assays for multiple cell variants are equivalent (i.e., poorly metastatic clones in the spontaneous metastasis assay are poorly metastatic in experimental metastasis assays. Likewise, individual clones are highly metastatic in both assays (20)). For experimental metastasis assays, cells ( $2 \times 10^5$  to  $3 \times 10^5$ ) suspended in ice-cold Hanks' balanced salt solution (0.2 mL) were injected into the lateral tail vein of 3- to 4-week-old female athymic nude mice. After 4 weeks, the mice were killed by cervical dislocation, and metastatic lesions were scored as described (20,24). Quantification of metastasis formation was identical, regardless of the metastasis assay used. Briefly, the lungs from each mouse were removed, fixed in a mixture of formalin and Bouin's fixative (5:1 vol/vol), and examined under a dissecting stereomicroscope (24). Unless otherwise noted, all other tissues were examined and found to be free of metastases.

### Statistical Analysis

The number of lung metastases was compared in KiSS-1 transfectants and parental C8161 cells. For experimental metastasis assays, one-way analysis of variance (ANOVA) followed by Tukey's Honestly Significant Difference post-test was used. For spontaneous metastasis assays, a Kruskal-Wallis ANOVA of ranks procedure was used. Calculations were performed using SigmaStat statistical analysis software (Jandel Scientific, San Rafael, CA). Statistical significance was defined as  $P \leq 0.05$  using two-sided tests.

## Results and Discussion

Several candidate metastasis-suppressor cDNAs were identified by subtractive hybridization and differential display comparing C8161 and neo6/C8161.1 cells (25). One clone, designated KiSS-1, was expressed only in nonmetastatic neo6/C8161.1 cells. The cDNA designation combines interim laboratory nomenclature for putative Suppressor Sequences with acknowledgment of the gene's discovery in Hershey. KiSS-1 cDNA sequence was submitted to GenBank as a novel gene with an accession number of U43527.

### Lack of Expression of KiSS-1 in Metastatic Melanoma Cells

The expression pattern of KiSS-1 was confirmed by northern blot analysis using an extensive panel of cell lines derived from the C8161 melanoma with widely ranging metastatic potentials. KiSS-1 mRNA expression could not be detected in any meta-

static melanoma cell line (Fig. 1, A). Expression was not detectable, even in overexposed blots or by RT-PCR, in parental C8161 or two subclones, C8161cl.9 and C8161cl.8, representing the highest and lowest metastatic potential among clones, respectively (12). KiSS-1 mRNA expression was also undetectable in C8161cl.9 microcell hybrids containing human chromosome 6 bearing a complex deletion within the region 6q21→q23 that remained metastatic in nude mice (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR: manuscript in preparation).

Seven independently prepared hybrids that contained an intact copy of human chromosome 6 in C8161 expressed high levels of a 1.0-kb transcript of KiSS-1 mRNA. Equal loading of lanes was confirmed by measuring glyceraldehyde-3-phosphate dehydrogenase (GAPDH) expression. All neo6/C8161 hybrid clones failed to metastasize following intravenous (experimental metastasis assay), subcutaneous, or intradermal (spontaneous metastasis assay) injection into 3- to 4-week-old female athymic nude mice. Therefore, KiSS-1 mRNA expression appeared to be a qualitative marker—i.e., exclusive to nonmetastatic C8161 cell populations.

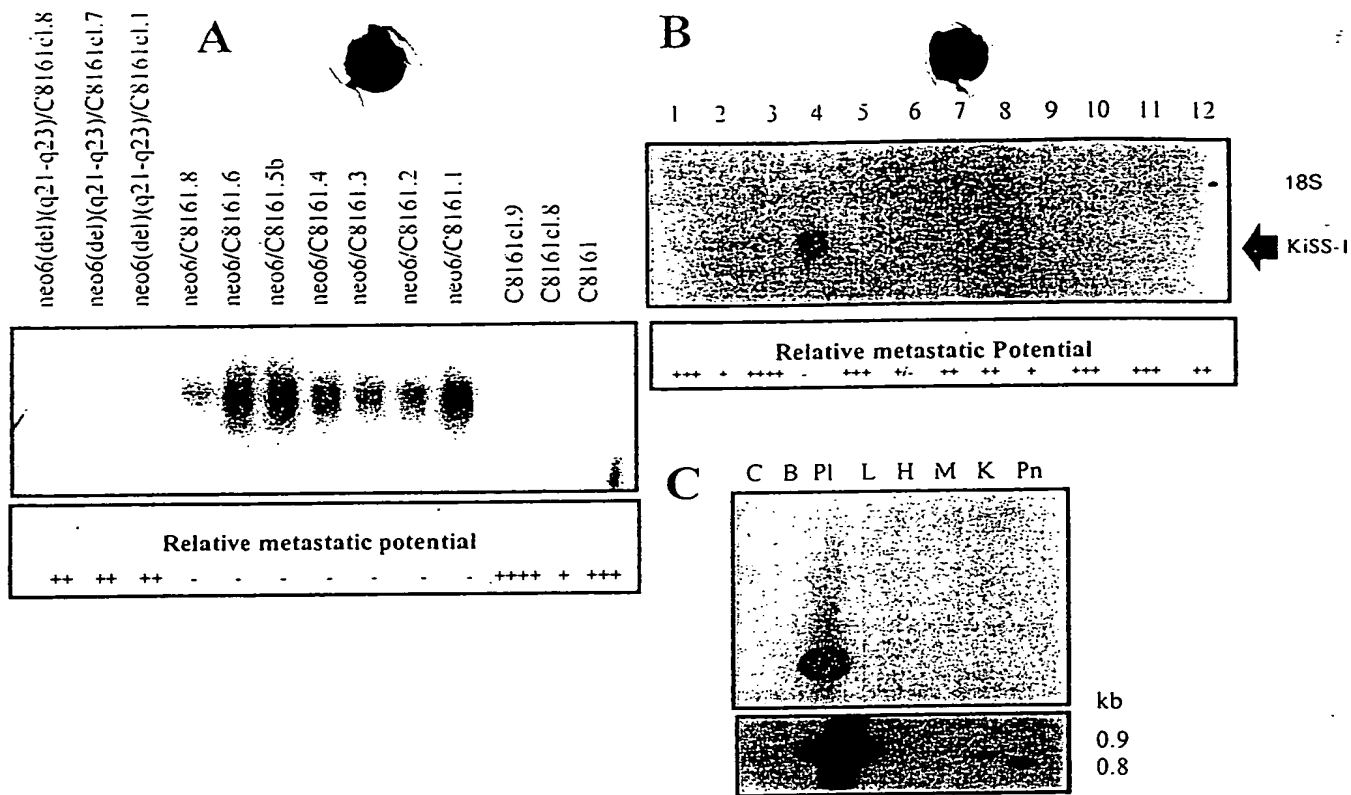
Expression of KiSS-1 mRNA was evaluated by northern blotting of other, unrelated human metastatic melanoma cell lines (Fig. 1, B). To test for the possibility that truncated forms of KiSS-1 may have existed, a full-length cDNA probe was used. We could not detect the 1.0-kb KiSS-1 transcript or bands of other sizes in any cells capable of metastasizing in athymic nude mice.

The expression of KiSS-1 mRNA in the human heart, brain, placenta, liver, lung, skeletal muscle, kidney, or pancreas was also examined by northern blot analyses (Fig. 1, C). Abundant KiSS-1 transcript (1.0 kb) was found in the placenta, with very weak expression in the kidney (detectable only after exposure for  $\geq 3$  days using two intensifying screens). Upon longer exposure, different transcript sizes of KiSS-1 were detected in the pancreas (0.8 kb) and the kidney (0.9 kb), suggesting that alternative splicing may take place in different tissues. Expression of KiSS-1 could also be detected in normal human melanocytes by RT-PCR (data not shown). Expression in melanocytes suggests that KiSS-1 functions in normal precursor cells and may therefore be useful in staging melanoma progression. The precise role of KiSS-1 in melanocyte function has not yet been determined.

KiSS-1 cDNA has a single open reading frame that encodes a protein of 164 amino acids with a predicted molecular mass of 18 kd (Fig. 2). The initiation codon and surrounding nucleotides fit the Kozak consensus, and there is a consensus polyadenylation site downstream of the termination codon. In vitro transcription and translation of a full-length KiSS-1 cDNA resulted in a single band at 18 kd (data not shown). A search of the GenBank and European Molecular Biology Laboratory databases revealed no significant homology with known genes. A similar lack of homology was seen for protein sequences in the SwissBank database.

### Metastasis Suppressed in C8161 KiSS-1 Transfectants

To demonstrate a functional basis for the correlation of KiSS-1 expression and the lack of metastatic potential, full-length



**Fig. 1.** Northern blot analyses of KiSS-1. **A**) Expression of KiSS-1 messenger RNA (mRNA) in cell clones derived from C8161, neo6/C8161, and neo6(del)(q21-q23)/C8161 cell clones. neo6/C8161 hybrids were prepared using MCH262A1.D6 as a microcell chromosome donor (27), and neo6(del)(q21-q23) hybrids were prepared using a related variant, A9/6q'Cl.1.2 (22), as a microcell chromosome donor (Miele ME, Goldberg SF, Hyatt DL, Barbanti-Brodano G, Welch DR; manuscript in preparation). Relative metastatic potentials are based on both experimental (intravenous inoculation into lateral tail vein) and spontaneous (intradermal inoculation into dorsolateral flank) metastasis assays. Poly(A)<sup>+</sup> mRNA (2.5 µg) was electrophoresed on denatured agarose gel, transferred onto a nylon membrane and fixed and probed with full-length KiSS-1 cDNA. KiSS-1 was undetectable in any cell line possessing metastatic potential, whether low or high. **B**) Expression of KiSS-1 mRNA in human metastatic melanoma cell lines. Lane 1 = C8161; 2 = C8161cl.8; 3 = C8161cl.9; 4 = non-

metastatic hybrid cell clone neo6/C8161.1; 5 = MelJuSo; 6 = A375P; 7 = A375M; 8 = MeWo; 9 = MeWo-3SS; 10 = MeWo-70W; 11 = M24met; and 12 = OM431. Poly(A)<sup>+</sup> RNA was isolated from each cell culture, and northern blot analysis was performed as above. All cell lines having metastatic potential did not express KiSS-1. **C**) Expression of KiSS-1 mRNA in normal human tissues. The multiple-tissue RNA blot [2 µg of poly(A)<sup>+</sup> RNA per lane] was purchased by use of a full-length KiSS-1 probe as above. Detection of KiSS-1 message (1.0 kilobase [kb]) in the placenta was possible following overnight exposure. KiSS-1 mRNA was detectable in the pancreas (0.8 kb) and in the kidney (0.9 kb) only after a 3-7-day exposure at -70 °C with two intensifying screens. Besides differential organ expression, these results suggest that alternative splice variants of KiSS-1 may exist in different cells.

cDNA of KiSS-1 was subcloned into the pcDNA3 constitutive expression vector and transfected into C8161. The vector alone was transfected as a negative control. Several clones were randomly selected and analyzed for expression of KiSS-1 by RNA blotting (Fig. 3). The KiSS-1 transcript in transfectants is larger (1.3 kb) because of additional sequences in the expression vector. The in vitro growth rates of the selected clones were not significantly ( $P > .05$ ) different compared with parental metastatic C8161 cell clones (Fig. 4). Clones expressing differing levels of KiSS-1 transcript were tested for metastatic ability in athymic nude mice by use of the experimental and spontaneous metastasis assays.

In the spontaneous metastasis assay, which measures the ability of cells injected intradermally or subcutaneously into the dorsolateral flank to metastasize to distant sites (20,24), KiSS-1 transfectant clones 2, 3, and 9 were less able to colonize lung or

regional lymph nodes than concomitantly injected C8161 cells (Table 1). Differences in metastatic potential could not be explained by slower in vivo primary tumor growth rate, since several more metastatic subclones of C8161 grow more slowly than KiSS-1 transfectants (12). Parental C8161 cells yielded an average of 50 lung metastases per mouse, and every mouse had regional lymph node metastases. In contrast, transfectant clone KiSS-1cl.2, which highly expressed KiSS-1 RNA, produced an average of only one metastasis per mouse. None of the mice had more than three metastases, and only two had lymph node metastases. KiSS-1 cl.3 produced a mean of three metastases per mouse (all mice yielded less than six metastases) and slightly higher incidences of lymph node metastases. Similar results were obtained following direct inoculation into the lateral tail veins of athymic nude mice (experimental metastasis assay, Table 2).



Table 1. Suppression of spontaneous metastasis of C8161 human melanoma cells by KiSS-1\*

Cell line	Relative KiSS-1 expression	No. of mice with metastases/total No. of mice	Lung		Extrapulmonary
			Mean $\pm$ SEM	Median (range)	
C8161	-	16/16	50 $\pm$ 25	9.5 (2 to >200)	8/8
KiSS-1cl.2	+++	5/7	1.1 $\pm$ 0.4	1 (0 to 3)	2/8
KiSS-1cl.3	+	6/7	2.7 $\pm$ 0.7	3 (0 to 6)	4/8
KiSS-1cl.9	++	0/7	0	0	2/7

\*Cells ( $1 \times 10^6$ ) were injected intradermally into the dorsolateral flank of 3- to 4-week-old female athymic nude mice (seven to eight mice per group, Harlan Sprague-Dawley). When the mean tumor diameter (square root of the product of orthogonal measurements) reached 1.5-2.0 cm, mice were killed. All organs were examined for the presence of metastases, and many were confirmed by microscopic examination of hematoxylin-eosin-stained paraffin-embedded sections (4-6  $\mu$ m). Macroscopic lung metastases were quantified after being stained with a mixture of neutral-buffered formalin and Bouin's fixative (5:1) and were counted with the aid of a dissecting microscope as previously described (20,24). SEM = standard error of the mean.

interactions that are important for intracellular signaling and cytoskeletal organization (36-43). Many of these interactions involving SH3 domains have been directly or indirectly associated with various steps in the metastatic cascade. Therefore, KiSS-1 could suppress metastasis by regulating key signaling pathways important to one or more of these steps. KiSS-1 transfectants did not display significantly ( $P > .05$ ) influenced adhesion to the extracellular matrix components type-I collagen or fibronectin nor was adhesion to a complex basement membrane-like mixture, Matrigel, affected (data not shown).

KiSS-1 maps to chromosome 1q32-q41 by fluorescence in situ hybridization (data not shown). Its mapping to another location within the genome compels the hypothesis that KiSS-1 is regulated by the gene(s) on chromosome 6. Our data also suggest that KiSS-1 is an important downstream effector of a gene(s) encoded on chromosome 6 and are consistent with metastasis suppression following introduction of chromosome 1 into MelJuSo melanoma cells (13). This hypothesis can be further refined because of data failing to detect KiSS-1 mRNA in metastatic neo6(del)(q21-q23)/C8161 hybrids. The latter can be explained by: 1) KiSS-1 mutations, 2) KiSS-1 deletions in subclones of C8161, or 3) mutation or deletion of regulatory

gene(s) mapping to 6q21-q23. These hypotheses will require additional studies. The latter explanation is consistent with high-frequency loss of heterozygosity observed in late stage melanomas (44-46).

In summary, we cloned a new human melanoma metastasis-suppressor gene by subtractive hybridization comparing metastatic C8161 cells and nonmetastatic neo6/C8161.1 cells. Loss of expression correlates with metastatic potential in human melanoma cells, and its expression results in significant suppression of melanoma metastasis in athymic nude mice. It will be necessary to further analyze the KiSS-1 gene in melanoma progression. The predicted KiSS-1 protein has a proline-rich region with homology to SH3-binding domain that could predict a mechanism for melanoma metastasis suppression. KiSS-1 may function in cell signaling and/or cytoskeletal organization on the basis of the usual role of SH3 domains in these processes; however, there is no evidence directly supporting these hypotheses. Taken together, these data strongly support a critical role for KiSS-1 in the regulation of cancer metastasis in human malignant melanoma. Additional experiments will be required to determine a role for KiSS-1 in other cancers or whether KiSS-1 may be a useful marker for staging melanomas in a clinical setting.

Table 2. Suppression of experimental metastasis of C8161 human melanoma cells by KiSS-1\*

Cell line	Relative KiSS-1 expression	No. of mice with lung metastases/total No. of mice	Lung metastases		P
			Mean $\pm$ SEM	Median (range)	
C8161	-	12/12	152 $\pm$ 25	>200 (2 to >200)	
C8161cl.9	-	4/4	200 $\pm$ 41	>200 (98 to >200)	NSD
C8161-neomix	-	4/4	160 $\pm$ 26	174 (93 to >200)	NSD
KiSS-1cl.12	+	8/8	81 $\pm$ 31	40 (1 to >200)	$P < .05$
KiSS-1cl.3	+	16/16	82 $\pm$ 49	101 (2 to >200)	$P < .05$
KiSS-1cl.9	++	15/16	21 $\pm$ 13	24.5 (0 to 80)	$P < .001$
KiSS-1cl.10	+++	4/12	1 $\pm$ 1	0 (0 to 4)	$P < .0001$
KiSS-1cl.2	+++	9/16	0.9 $\pm$ 0.4	0.5 (0 to 3)	$P < .001$

\*Cells ( $3 \times 10^5$ ) were injected intravenously into the lateral tail vein of 3- to 4-week-old female athymic nude mice (seven to eight mice per group, Harlan Sprague-Dawley). Mice were killed 26 days after inoculation. All organs were examined for the presence of metastases, and many were confirmed by microscopic examination of hematoxylin-eosin-stained paraffin-embedded sections (4-6  $\mu$ m). Macroscopic lung metastases were quantified with the aid of a dissecting microscope after staining with a mixture of neutral-buffered formalin and Bouin's fixative (5:1) as described (20,24). P values were calculated by one-way analysis of variance using Tukey's Honestly Significant Difference post-test. Values shown compare each group with parental C8161. SEM = standard error of the mean; NSD = no significant difference.

## References

- (1) Liotta LA, Stetler-Stevenson WG. Tumor progression and metastasis: an imbalance of positive and negative regulation. *Cancer Res* 1991;51(18 Suppl):5054s-9s.
- (2) Liotta LA, Steeg PS, Stetler-Stevenson WG. Cancer metastasis and angiogenesis: an imbalance of positive and negative regulation. *Cell* 1991; 64:327-36.
- (3) Fidler IJ, Ellis LM. The implications of angiogenesis for the biology and therapy of cancer metastasis [see comment citation in Medline]. *Cell* 1994; 79:185-8.
- (4) Nicolson GL. Cancer progression and growth: relationship of paracrine and autocrine growth mechanisms to organ preference of metastasis. *Exp Cell Res* 1993;204:171-80.
- (5) Dear TN, Kefford RF. Molecular oncogenetics of metastasis. *Mol Aspects Med* 1990;11:243-324.
- (6) Wright JA, Egan SE, Greenberg AH. Genetic regulation of metastatic progression. *Anticancer Res* 1990;10:1247-55.
- (7) MacDougall JR, Matrisian LM. Contributions of tumor and stromal matrix metalloproteinases to tumor progression, invasion and metastasis. *Cancer Metastasis Rev* 1995;14:351-62.
- (8) Fodstad O, Kjonniksen I. Microenvironment revisited: time for reappraisal of some prevailing concepts of cancer metastasis. *J Cell Biochem* 1994; 56:23-8.
- (9) Ponta H, Hofmann M, Herrlich P. Recent advances in the genetics of metastasis. *Eur J Cancer* 1994;30A:1995-2001.
- (10) Degen WG, Weterman MA, van Groningen JJ, Cornelissen IM, Lemmers JP, Agterbos MA, et al. Expression of nma, a novel gene, inversely correlates with the metastatic potential of human melanoma cell lines and xenografts. *Int J Cancer* 1996;65:460-5.
- (11) Safarians S, Sternlicht MD, Freiman CJ, Huaman JA, Barsky SH. The primary tumor is the primary source of metastasis in a human melanoma/scid model. Implications for the direct autocrine and paracrine epigenetic regulation of the metastatic process. *Int J Cancer* 1996;66:151-8.
- (12) Welch DR, Chen P, Miele ME, McGary CT, Bower JM, Stanbridge EJ, et al. Microcell-mediated transfer of chromosome 6 into metastatic human C8161 melanoma cells suppresses metastasis but does not inhibit tumorigenicity. *Oncogene* 1994;9:255-62.
- (13) Miele ME, Robertson G, Lee JH, Coleman A, McGary CT, Fisher PB, et al. Metastasis suppressed, but tumorigenicity and local invasiveness unaffected, in the human melanoma cell line Me/JuSo after introduction of human chromosomes 1 or 6. *Mol Carcinog* 1996;15:284-99.
- (14) You J, Miele ME, Dong C, Welch DR. Suppression of human melanoma metastasis by introduction of chromosome 6 may be partially due to inhibition of motility, but not to inhibition of invasion. *Biochem Biophys Res Commun* 1995;208:476-84.
- (15) Liang P, Pardee AB. Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction [see comment citation in Medline]. *Science* 1992;257:967-71.
- (16) Liang P, Averboukh L, Pardee AB. Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucleic Acids Res* 1993;21:3269-75.
- (17) van Groningen JJ, Bloemers HP, Swart GW. Identification of melanoma inhibitory activity and other differentially expressed messenger RNAs in human melanoma cell lines with different metastatic capacity by messenger RNA differential display. *Cancer Res* 1995;55:6237-43.
- (18) Lee SW, Tomasello C, Sager R. Positive selection of candidate tumor-suppressor genes by subtractive hybridization. *Proc Natl Acad Sci U S A* 1991;88:2825-9.
- (19) Duguid JR, Rohwer RG, Seed B. Isolation of cDNAs of scrapie-modulated RNAs by subtractive hybridization of a cDNA library. *Proc Natl Acad Sci U S A* 1988;85:5738-42.
- (20) Welch DR, Bisi JE, Miller BE, Conaway D, Seftor EA, Yohem KH, et al. Characterization of a highly invasive and spontaneously metastatic human malignant melanoma cell line. *Int J Cancer* 1991;47:227-37.
- (21) Trent JM, Stanbridge EJ, McBride HL, Meese EU, Casey G, Araujo DE, et al. Tumorigenicity in human melanoma cell lines controlled by introduction of human chromosome 6. *Science* 1990;247:568-71.
- (22) Gualandri F, Morelli C, Pavan JV, Rimessi P, Senzi A, Bonfatti A, et al. Induction of senescence and control of tumorigenicity in BK virus transformed mouse cells by human chromosome 6. *Genes Chromosomes Cancer* 1994;10:77-84.
- (23) Schweinfest CW, Henderson KW, Gu JR, Kottaridis SD, Besbas S, Panotopoulou E, et al. Subtraction hybridization cDNA libraries from colon carcinoma and hepatic cancer. *Genet Anal Tech Appl* 1990;7:64-70.
- (24) Welch DR, Neri A, Nicolson GL. Comparison of spontaneous and 'experimental' metastasis using rat 13762 mammary adenocarcinoma cell clones. *Invasion Metastasis* 1983;3:65-80.
- (25) Lee JH, Welch DR. Differential display and subtractive hybridization to identify highly expressed genes in metastasis-suppressed chromosome 6/human melanoma hybrids. *Proc Am Assoc Cancer Res* 1996; 37:530.
- (26) Cohen GB, Ren R, Baltimore D. Modular binding domains in signal transduction proteins. *Cell* 1995;80:237-48.
- (27) Alexandropoulos K, Cheng G, Baltimore D. Proline-rich sequences that bind to Src homology 3 domains with individual specificities. *Proc Natl Acad Sci U S A* 1995;92:3110-4.
- (28) von Heijne G. A new method for predicting signal sequence cleavage sites. *Nucleic Acids Res* 1986;14:4683-90.
- (29) Hennessey C, Henry JA, May FE, Westley BR, Angus B, Lennard TW. Expression of the antimetastatic gene nm23 in human breast cancer: an association with good prognosis. *J Natl Cancer Inst* 1991;83:281-5.
- (30) Musacchio A, Wilmanns M, Saraste M. Structure and function of the SH3 domain. *Prog Biophys Mol Biol* 1994;61:283-97.
- (31) Joseph G, Pick E. Peptide walking is a novel method for mapping functional domains in proteins. Its application to the Rac1-dependent activation of NADPH oxidase. *J Biol Chem* 1995;270:29079-82.
- (32) Feng S, Chen JK, Yu H, Simon JA, Schreiber SL. Two binding orientations for peptides to the Src SH3 domain: development of a general model for SH3-ligand interactions. *Science* 1994;266:1241-7.
- (33) Mayer BJ, Eck MJ. SH3 domains. Minding your p's and q's. *Curr Biol* 1995;5:364-7.
- (34) Rickles RJ, Botfield MC, Zhou XM, Henry PA, Brugge JS, Zoller MJ. Phage display selection of ligand residues important for Src homology 3 domain binding specificity. *Proc Natl Acad Sci U S A* 1995;92:10909-13.
- (35) Ishino M, Ohba T, Sasaki H, Sasaki T. Molecular cloning of a cDNA encoding a phosphoprotein, Efs, which contains a Src homology 3 domain and associates with Fyn. *Oncogene* 1995;11:2331-8.
- (36) Koch CA, Anderson D, Moran MF, Ellis C, Pawson T. SH2 and SH3 domains: elements that control interactions of cytoplasmic signaling proteins. *Science* 1991;252:668-74.
- (37) Musacchio A, Noble M, Pauptit R, Wierenga R, Saraste M. Crystal structure of a Src-homology 3 (SH3) domain [see comment citation in Medline]. *Nature* 1992;359:851-5.
- (38) Pawson T, Gish GD. SH2 and SH3 domains: from structure to function. *Cell* 1992;71:359-62.
- (39) Pawson T, Schlessinger J. SH2 and SH3 domains. *Curr Biol* 1993;3:434-42.
- (40) Feller SM, Ren R, Hanafusa H, Baltimore D. SH2 and SH3 domains as molecular adhesives: the interactions of Crk and Abl. *TIBS Trends Biochem Sci* 1994;19:453-8.
- (41) Nobes CD, Hall A. Rho, Rac, and cdc42 GTPases regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia, and filopodia. *Cell* 1995;81:53-62.
- (42) Ridley AJ. Membrane ruffling and signal transduction. *Bioessays* 1994; 16:321-7.
- (43) Cantley LC, Auger KR, Carpenter C, Duckworth B, Graziani A, Kapeller R, et al. Oncogenes and signal transduction [published erratum appears in *Cell* 1991;65:following 914]. *Cell* 1991;64:281-302.
- (44) Trent JM, Thompson FH, Meyskens FL Jr. Identification of a recurring translocation site involving chromosome 6 in human malignant melanoma. *Cancer Res* 1989;49:420-3.
- (45) Guan XY, Meltzer PS, Cao J, Trent JM. Rapid generation of region-specific genomic clones by chromosome microdissection: isolation of DNA from a region frequently deleted in malignant melanoma. *Genomics* 1992;14:680-4.
- (46) Millikin D, Meese E, Vogelstein B, Witkowski C, Trent JM. Loss of heterozygosity for loci on the long arm of chromosome 6 in human malignant melanoma. *Cancer Res* 1991;51:5449-53.

## Notes

*Present address:* M. E. Miele, Department of Medical Technology, University of Delaware, Newark.

Supported by Public Health Service grants CA62168 and CA44470 from the National Cancer Institute, National Institutes of Health, Department of Health and Human Services, by the National Foundation for Cancer Research, and by gifts from the Jake Gittlen Memorial Golf Tournament.

This paper is dedicated to the memory of Ray C. Welch, who was always supportive and encouraging.

Manuscript received May 10, 1996; revised August 14, 1996; accepted September 10, 1996.